TRANSACTIONS

OF THE

AMERICAN INSTITUTE

OF

ELECTRICAL ENGINEERS

VOLUME 79

1960

PART I. COMMUNICATION AND ELECTRONICS

PUBLISHED BY THE
AMERICAN INSTITUTE OF ELECTRICAL ENGINEERS
33 WEST 39TH STREET
NEW YORK 18, N. Y., U. S. A.

Copyright 1961 by the American Institute of Electrical Engineers

Printed in U. S. A. by The Mack Printing Company, Easton, Pa.

Preface

The AIEE *Transactions* for 1960 (volume 79) is published in three parts: Part I. Communication and Electronics; Part II. Applications and Industry; and Part III. Power Apparatus and Systems. The papers in each of the three parts are classified according to subject matter as follows:

Part I. Communication and Electronics

Aero-Space Instrumentation
Basic Sciences
Broadcasting
Communication Switching Systems
Communication Theory
Computing Devices
Data Communication
Electrical Insulation
Electrical Techniques in Medicine and Biology
Electronic and High-Frequency Instruments
Electronics
Fundamental Electrical Standards
Indicating and Integrating Instruments
Magnetic Amplifiers
Nuclear Instrumentation
Nucleonics
Radio Communication Systems
Recording and Controlling Instrumentation
Semiconductor Rectifiers
Solid State Devices
Space Communications
Special Instruments and Auxiliary Apparatus
Telegraph Systems
Telemetering
Wire Communication Systems

Part II. Applications and Industry

Aero-Space Transportation
Chemical Industry
Domestic and Commercial Applications
Electric Heating
Electric Welding
Feedback Control Systems
General Industry Applications
Industrial and Commercial Power Systems
Industrial Control
Industrial Power Rectifiers
Land Transportation
Marine Transportation
Machine Tools Industry
Metal Industry
Mining Industry
Petroleum Industry
Production and Application of Light
Rubber and Plastic Industries
Textile Industry

Part III. Power Apparatus and Systems

Insulated Conductors
Power Generation
Power System Communications
Protective Devices
Relays
Rotating Machinery
Substations
Switchgear
System Engineering
Transformers
Transmission and Distribution

New Technical Activities
Research
Safety
Standards

Each part has been indexed separately in the back of that particular part. The three parts are not cross-referenced; hence the user should determine first whether the subject matter of the paper desired is in the field of communication and electronics, applications and industry, or power apparatus and systems. Papers are listed in the subject index under several key words in the titles. The original numbers assigned to the papers are given in the author index. Volume 79 contains the technical papers and related discussions presented at these meetings:

1. Winter General Meeting, New York, N. Y., January 31–February 5, 1960.
2. Southwest District Meeting, Houston, Tex., April 4–6, 1960.
3. Great Lakes District Meeting, Milwaukee, Wis., April 27–29, 1960.
4. Northeastern District Meeting, Providence, R. I., May 2–4, 1960.
5. Summer General Meeting, Atlantic City, N. J., June 19–24, 1960.
6. Pacific General Meeting, San Diego, Calif., August 8–12, 1960.
7. Fall General Meeting, Chicago, Ill., October 9–14, 1960.

Statements and opinions given in the papers and discussions published in *Transactions* are the expressions of the contributors for which the American Institute of Electrical Engineers assumes no responsibility.

Communication and Electronics Index for 1960

1. 79
1
60

1. Technical Subject Index

A-C Powered Cranes, D-C Magnetic Crane Hoist Control for. Myles, Davies, Srnka........207–11
Accurate Voltage Measurements to 30 Megacycles Per Second, Thermal Voltage Converters for. Hermach, Williams................200–05; disc. 205
Air-Borne Ice and Snow Measurement, Proposed Monocycle-Pulse Very-High-Frequency Radar for. Cook........................588–93; disc. 593
Air-Borne Real-Time Control, The Relative Merits of Incremental and Conventional Digital Computers in. Shackell, Tryon....................393–400
Air-Cooled Nuclear Reactors, A Digital Start-up Control for. Lehr, Mathis...............369–75
Air Ions on the Trachea, Experiments with. Beckett, Krueger.................................695–9
Air, The Effects of Nuclear Radiation on the Electrical Strength of. Duncan, Fraser, Valachovic...19–26
Aircraft Nuclear Reactor, Magnetic Automatic Power-Range Control for an. Russell, Hemmenway, Scharf, Sharr...........................379–84
Airlines Reservations, Teletypewriter Selective Calling Applied to Automatic. Bratschi, Pennington, Whitman..............................579–83
All-Transistorized Trunk Carrier System, An. Gardner 212–16
Amplidyne Circuit Analysis. Bills.............711–15
Amplifier, A High-Speed Microcurrent. Dever, Sickles 375–9
Amplifiers, Stable Transistor Wide-Band D-C. Okada 26–33
Analog-Digital Simulation, A 2-Channel Data Link for Combined. Greenstein........................40–4
Analog of a Magnetic Circuit Established by the Basic Laws Defining Magnetostatics, The Existence of an Electric. Koenig....................400–03
Analog Solution of Some Functional Analysis Problems, The. Graupe..................793–7; disc. 798
Analog with Spiral Search Coils, The Magnetic. Moore ....256–63; disc. 263
Analysis, Amplidyne Circuit. Bills............711–15
Analysis and Optimization of the Bridge Magnetic Servoamplifier with True Motor Load. Walsh, Meiksin..................................413–20
Analysis of a Polarized Electromagnetic Drive System and of Vibrator Drive Systems, An. Nunn..346–52
Analysis of Self-Saturating Magnetic Amplifiers, Rectifier Unblocking from a Generalized. Bourne, Nitzan...................................99–105
Analysis of the Operation of the Magnetic Second-Harmonic Modulator, An. Jalbert.......268–72
Analysis of Transcendental Nonlinear Systems. Wolf 449–51
Analyzer and General Purpose Equipment in Guidance Systems, A Comparison of Digital Differential. Dickinson..............................706–08
Antenna Array Excited Parasitically by a Central Driven Element, An Omnidirectional Circular. Neff, Tillman..................................190–2
Applications, A New Ceramic Triode for VHF. Campbell.....................................404–08
Approach to Dynamic Optimizing Control of the Continuous Process, An. Sandelien........291–9
Arc Extinction, Voltage Transients Due to. Steiner, Strecker...................................139–44
Arsenal Closed-Circuit Educational Television System, Redstone. Argo, Howell.................185–90
Assessment, Tubes or Transistors: A Realistic. Moe. 81–5
Assignment of Carry Variables in Iterative Networks. McCluskey..............................772–8
Audio Frequencies, A Transistor Automatic Gain Control System for. Shirman.............664–9
Audio Frequencies with 1-Microsecond Precision, A Test Set for Measurement of Envelope Delay Distortion at. Codd.......................241–5
Automatic Airlines Reservations, Teletypewriter Selective Calling Applied to. Bratschi, Pennington, Whitman..............................579–83
Automatic Gain Control System for Audio Frequencies, A Transistor. Shirman..................664–9
Automatic Power-Range Control for an Aircraft Nuclear Reactor, Magnetic. Russell, Hemmenway, Scharf, Sharr.......................379–84
Automation in Switching Equipment Engineering. Hacker, Davis...........................545–52
Automation of Computer Panel Wiring. Altman, DeCampo, Warburton...................118–25
Automation of Telephone Switching Equipment Engineering. Blumer, Reifenstuhl, List.........560–4

B

Back-Transient Diode Logic. Wolff..............4–9
Barium Titanate, Quantitative Study of Low-Frequency Hysteresis Loops of Polarized Polycrystalline. Marks, Hanna.........................799–808
Basic Concepts for Magnet Coil Design, Some. Gauster 822–8
Basic Laws Defining Magnetostatics, The Existence of an Electric Analog of a Magnetic Circuit Established by the. Koenig.........................400–03
(Basic Sciences) A Method for the Design of Holding Electromagnets. Ludwig..................300–10
(Basic Sciences) An Analysis of a Polarized Electromagnetic Drive System and of Vibrator Drive Systems. Nunn.........................346–52
(Basic Sciences) Analysis of Transcendental Nonlinear Systems. Wolf.........................449–51
(Basic Sciences) Difference Equations and Their Applications. Pipes.............333–9; disc. 339
(Basic Sciences) Generalized Brillouin Flow. Kent... 144–8
(Basic Sciences) Inductors Biased with Permanent Magnets. Ludwig:
  Part I.....................................273–8
  Part II....................................278–91
(Basic Sciences) On the Response of Linear Systems to Periodic Excitations. LePage, McFee.......... ....746–8; disc. 749
(Basic Sciences) Quantitative Study of Low-Frequency Hysteresis Loops of Polarized Polycrystalline Barium Titanate. Marks, Hanna........799–808
(Basic Sciences) The Properties of Pole and Zero Locations for Nondecreasing Step Responses. Zemanian................................421–6
(Basic Sciences) The Synthesis of Linear Filters with Real or Imaginary Transfer Functions. Gilbert, Otterman................................323–30
Binary Operations, Switching Function Canonical Forms Based on Commutative and Associative. Calingaert...............................808–14
Biphase Rectifiers, Saturable-Reactor Control of Full-Wave and. Thomas, Drenning.............36–9
Bistable Transistor Pulse Circuits, Fundamental Considerations of Power Dissipation Limits in Some. Raillard..................................53–5
Bistatic Radar at 3,000 Mc, Moon Reflection Studies with. Straiton, Tolbert........436–9; disc. 439
Brake for a Sodium-Cooled Nuclear Power Reactor, Design of an Eddy-Current. Baker.........330–3
Breakdown Phenomenon in Selenium Rectifiers, A Nondestructive. English, Tobin.............9–14
Bridge Magnetic Amplifier, Theory of 3-Phase. Storm, Flairty:
  Part II....................................451–8
  (Part I—1959)
Bridge Magnetic Servoamplifier with True Motor Load, Analysis and Optimization of the. Walsh, Meiksin 413–20
Brillouin Flow, Generalized. Kent.............144–8
(Broadcasting) An Omnidirectional Circular Antenna Array Excited Parasitically by a Central Driven Element. Neff, Tillman..................190–2
(Broadcasting) Dial-Selected Industrial Television for Ticket Reservation Facilities. Alinsky......174–7
(Broadcasting) New Television Network Switching Facilities. Bachelet, Collins, Taylor.......625–34
(Broadcasting) Production Design Considerations in Television Chassis Sweep Interference Reduction. Carson...................................557–60
(Broadcasting) Redstone Arsenal Closed-Circuit Educational Television System. Argo, Howell...185–90

C

Cable Carrier Repeater, A Transistorized. Griffith.. 675–80
Calling Applied to Automatic Airlines Reservations, Teletypewriter Selective. Bratschi, Pennington, Whitman...................................579–83
Canonical Forms Based on Commutative and Associative Binary Operations, Switching Function. Calingaert...............................808–14
Carrier Communications on the Union Pacific Railroad, Wire Line and. Jett......................520–5
Carrier Repeater, A Transistorized Cable. Griffith.. 675–80
Carrier System, An All-Transistorized Trunk. Gardner 212–16
Carrier Systems, Two-Wire Operation of N. Dean, Karkar....................................594–7
Carry Variables in Iterative Networks, Assignment of. McCluskey..............................772–8
Centers, Voice Immunity of Tone Operating Switching. Guenther, Shirman...........................553–7
Central Driven Element, An Omnidirectional Circular Antenna Array Excited Parasitically by a. Neff, Tillman.................................190–2
Centralized Testing of a Very-High-Frequency Mobile Radio System. MacDonald................641–5
Ceramic Triode for VHF Applications, A New. Campbell.....................................404–08
Channel Reactor Protection System for Nuclear Power Plants, A Dual-. Bartu...................358–62
Channels for Doubling TD-2 Radio System Capacity, Interstitial. Curtis, Collins, Jamison......646–54
Character-Metered Transatlantic Switching System, A. Coggeshall, Holcomb............56–64; disc. 64
Characteristics of a High-Speed Wide-Chart Recorder, Design and Performance. Garrigus.......163–70
Charge-Limited Currents with Application to Electrical Precipitation, A Theory for Space-. Cooperman. 47–50
Charges on the Inception Voltage Between Dielectric and Metal Dielectric Surfaces, The Measurement and Influence of Surface. Reynolds.......310–14
Chemical Processing Model for the Investigation of Computer Control, A Generalized. Williams, Otto.....................................458–73
Chemical Reactor, Dynamic Optimization and Control of a Stirred-Tank. Gould, Kipiniak.......... ....734–43; disc. 743
Circuit Analysis, Amplidyne. Bills............711–15
Circuit Established by the Basic Laws Defining Magnetostatics, The Existence of an Electric Analog of a Magnetic. Koenig.......................400–03
Circuits, Fundamental Considerations of Power Dissipation Limits in Some Bistable Transistor Pulse. Raillard...................................53–5
Circuits, Symmetrical Frequency Multiplier. Dick.. 125–34
Circular Antenna Array Excited Parasitically by a Central Driven Element, An Omnidirectional. Neff, Tillman..............................190–2
Closed-Circuit Educational Television System, Redstone Arsenal. Argo, Howell.................185–90
Codes for Correcting Bursts of Errors, Error-Correcting. Meggitt.............................708–11
Coil Design, Some Basic Concepts for Magnet. Gauster 822–8
Coils, The Magnetic Analog with Spiral Search. Moore ....256–63; disc. 263
Combi-System—A Proposal for New Concepts in Digital Data Processing, The. Schwab............193–7
(Committee Report) Proposed Symbology for Digital Systems. (60–1224)............787–93; disc. 793
(Communication Switching Systems) A 240-Line Fully Electronic Telephone Switchboard. Goudet..... 232–41
(Communication Switching Systems) Automation in Switching Equipment Engineering. Hacker, Davis 545–52
(Communication Switching Systems) Automation of Telephone Switching Equipment Engineering. Blumer, Reifenstuhl, List..................560–4
(Communication Switching Systems) Radio-Frequency Interference Considerations in the TD-2 Radio Relay System. Curtis.....................64–70
(Communication Switching Systems) Reflections on the Traffic Usage Recorder with Emphasis on Traffic Engineering. Bauerfeld...................566–70
(Communication Switching Systems) Synthesis of a Resonant Transfer Filter as Applied to a Time Division Multiplex System. May, Stump..615–20
(Communication Switching Systems) The New Nation-Wide Telephone Numbering Plan. Myers..673–5
(Communication Switching Systems) Voice Immunity of Tone Operated Switching Centers. Guenther, Shirman...................................553–7
(Communication Theory) On the Transmission of Information by Orthogonal Time Functions. Harmuth...................................248–55
Communication with Orthogonal Time Functions, Radio. Harmuth.........................221–8

Vol.
Pt. 1
196

Communications on the Union Pacific Railroad, Wire Line and Carrier. Jett....................520–5
Communications Systems, Data Processing as a Tool for Generalizing. Luebbert....................431–6
Communications, Use of Moon or Satellite Relays for Global. Yeh............................607–15
Commutative and Associative Binary Operations, Switching Function Canonical Forms Based on. Calingaert....................................808–14
Comparison of Digital Differential Analyzer and General Purpose Equipment in Guidance Systems, A. Dickinson....................................706–08
Computer Control, A Generalized Chemical Processing Model for the Investigation of. Williams, Otto..................................458–73
Computer Control System, The LN 3000. Harple, Lex....................................408–13
Computer Panel Wiring, Automation of. Altman, DeCampo, Warburton....................118–25
Computer Use, Specifying a Pulse Transformer for. Blessing....................................44–7
Computers at the Technical University of Denmark, Teaching the Uses of Digital. Franksen....781–7
Computers, Electronic Transformer Design by Digital. Deise, Etchison, Lee............314–22; disc. 322
Computers for Process Control, Hybrid. Birkel..................................726–34; disc. 734
Computers in Air-Borne Real-Time Control, The Relative Merits of Incremental and Conventional Digital. Shackell, Tryon....................393–400
(Computing Devices) A Comparison of Digital Differential Analyzer and General Purpose Equipment in Guidance Systems. Dickinson............706–08
(Computing Devices) A Magnetic-Tape-to-Paper-Tape Converter. Ringer, Mintzer........339–46
(Computing Devices) A New Function Generator. Eggers, Sze....................................756–62
(Computing Devices) A 2-Channel Data Link for Combined Analog-Digital Simulation. Greenstein...40–4
(Computing Devices) Amplidyne Circuit Analysis. Bills....................................711–15
(Computing Devices) An Approach to Dynamic Optimizing Control of the Continuous Process. Sandelien....................................291–9
(Computing Devices) Assignment of Carry Variables in Iterative Networks. McCluskey.........772–8
(Computing Devices) Current-Operated Diode Logic Gates. Reinecke........................762–72
(Computing Devices) Dynamic Optimization and Control of a Stirred-Tank Chemical Reactor. Gould, Kipiniak..............734–43; disc. 743
(Computing Devices) Error-Correcting Codes for Correcting Bursts of Errors. Meggitt.........708–11
(Computing Devices) Logical Organization of the Honeywell H-290. Eachus...............715–19
(Computing Devices) Magnetic Memory Drum Design. Carne....................................749–56
(Computing Devices) Proposed Symbology for Digital Systems. (Committee Report)..787–93; disc. 793
(Computing Devices) Switching Function Canonical Forms Based on Commutative and Associative Binary Operations. Calingaert...........808–14
(Computing Devices) The Analog Solution of Some Functional Analysis Problems. Graupe..............................793–7; disc. 798
(Computing Devices) The Magnetic Analog with Spiral Search Coils. Moore..........256–63; disc. 263
(Computing Devices) Universal Modulus—Matrix Logic VI. Schubert........................778–81
Concepts for Magnet Coil Design, Some Basic. Gauster....................................822–8
Constant and Transfer Function of Magnetic Amplifiers, Time Response. Manteuffel....473–82; disc. 482
Continuous Process, An Approach to Dynamic Optimizing Control of the. Sandelien.............291–9
Continuous Voltage Reference for Industrial Recorders, A Precision. Robinson.................197–200
Control for Air-Cooled Nuclear Reactors, A Digital Start-up. Lehr, Mathis....................369–75
Control for an Aircraft Nuclear Reactor, Magnetic Automatic Power-Range. Russell, Hemmenway, Sharf, Sharr....................................379–84
Control of Full-Wave and Biphase Rectifiers, Saturable-Reactor. Thomas, Drenning...............36–9
Control of the Continuous Process, An Approach to Dynamic Optimizing. Sandelien...........291–9
Control System, The LN 3000 Computer. Harple, Lex....................................408–13
Control Terminal for Mobile Telephone Service, A New. Getgen, McGibbon................583–8
Conventional Digital Computers in Air-Borne Real-Time Control, The Relative Merits of Incremental and. Shackell, Tyron....................393–400
Converter, A Magnetic-Tape-to-Paper-Tape. Ringer Mintzer....................................339–46
Converters for Accurate Voltage Measurements to 30 Megacycles Per Second, Thermal Voltage. Hermach, Williams................200–05; disc. 205
Cores, Magnetization in Tape-Wound. Barker..............................482–501; disc. 501
Corona, Current-Voltage Relationships in Negative Pulsed. Thomas, Williams................136–9
Corona Fields, Power Relationships and Temperature Dependence in the D-C. Thomas, Williams, Suzuki....................................1–4
Corona Fields, Potentials in D-C. Penney, Matick..91–9
Corona, Sparkover as Influenced by Surface Conditions in D-C. Penney, Craig....................112–18
Correcting Codes for Correcting Bursts of Errors, Error-. Meggitt....................................708–11
Count-Rate and Period Meters, Transients in Logarithmic. Barrow, Maitland....................362–9
Coupled Multivibrators, Recent Developments on Magnetic.- Geyger....................106–12
Crane Hoist Control for A-C Powered Cranes, D-C Magnetic. Myles, Davies, Srnka..........207–11
Crestatrons, Experiments on a Series of S-Band. Rowe, Konrad, Krage........................828–33
Cross-Field Magnetic Modulator, Theory and Operation of a. Boensel..................505–10; disc. 510
Current-Operated Diode Logic Gates. Reinecke..762–72
Current Transformers for Multirange Operation of Milliammeter Ink Recorders, Miniaturized. Geyger....................................681–9
Current-Voltage Relationships in Negative Pulsed Corona. Thomas, Williams................136–9
Currents with Application to Electrical Precipitation A Theory for Space-Charge-Limited. Cooperman..47–50
Cyclotrons, Sector-Focused. Blosser...........814–22

D

D-C Amplifiers, Stable Transistor Wide-Band. Okada..26–33
D-C Corona Field, Power Relationships and Temperature Dependence in the. Thomas, Williams, Suzuki....................................1–4
D-C Corona Fields, Potentials in. Penney, Matick..91–9
D-C Corona, Sparkover as Influenced by Surface Conditions in. Penney, Craig............112–18
D-C Magnetic Crane Hoist Control for A-C Powered Cranes. Myles, Davies, Srnka............207–11
D-C Supply, Static Control for a Mechanically Regulated. Abrams, Brubaker....................426–30
(Data Communication) A Test Set for Measurement of Envelope Delay Distortion at Audio Frequencies with 1-Microsecond Precision. Codd........241–5
(Data Communication) A Wired-Memory Translator with Shared Access. Pace, Ostendorf......216–20
Data Link for Combined Analog-Digital Simulation, A Two-Channel. Greenstein....................40–4
Data Processing as a Tool for Generalizing Communications Systems. Luebbert....................431–6
Data Processing, The Combi-System—A Proposal for New Concepts in Digital. Schwab.........193–7
Data Transmission on an Electric Power System, Development and Application of Time-Division Multiplexing Telemetering Equipment for. Almon, Donelson....................................155–9
Data Transmission, The Effect of Delay Distortion on. Mertz....................................228–32
Datacom Unit—An Experimental Data Transmission Subset. Stoffels....................................669–73
Delay Distortion at Audio Frequencies with 1-Microsecond Precision, A Test Set for Measurement of Envelope. Codd....................................241–5
Delay Distortion on Data Transmission, The Effect of. Mertz.................................... 228–32
Delay Distortion Simulation Set, A. Dunbar....183–5
Denmark, Teaching the Uses of Digital Computers at the Technical University of. Franksen.....781–7
Design and Performance Characteristics of a High-Speed Wide-Chart Recorder. Garrigus....163–70
Design Considerations for Radio Relays Utilizing the Tropospheric Scatter Mode of Propagation, Optimum. Parry....................................71–80
Design Considerations in Television Chassis Sweep Interference Reduction, Production. Carson..557–60
Design of an Eddy-Current Brake for a Sodium-Cooled Nuclear Power Reactor. Baker............330–3
Design of Holding Electromagnets, A Method for the. Ludwig....................................300–10
Development and Application of Time-Division Multiplexing Telemetering Equipment for Data Transmission on an Electric Power System. Almon, Donelson....................................155–9
Device Be Used? Shall an Electron Tube or a Semiconductor. Scheneman, Waldorf..........264–8
Dial-Selected Industrial Television for Ticket Reservation Facilities. Alinksy....................174–7
Dielectric and Metal Dielectric Surfaces, The Measurement and Influence of Surface Charges on the Inception Voltage Between. Reynolds....310–14
Difference Equations and Their Applications. Pipes..................................333–9; disc. 339
Digital Computers at the Technical University of Denmark, Teaching the Uses of. Franksen..781–7
Digital Computers, Electronic Transformer Design by. Deise, Etchison, Lee............314–22; disc. 322
Digital Computers in Air-Borne Real-Time Control, The Relative Merits of Incremental and Conventional. Shackell, Tryon....................393–400
Digital Data Processing, The Combi-System—A Proposal for New Concepts in. Schwab........193–7
Digital Differential Analyzer and General Purpose Equipment in Guidance Systems, A Comparison of. Dickinson....................................706–08
Digital Magnetic Recorder, A Step-Driven. Hansen..564–5
Digital Simulation, A 2-Channel Data Link for Combined Analog-. Greenstein....................40–4
Digital Start-up Control for Air-Cooled Nuclear Reactors, A. Lehr, Mathis....................369–75
Digital Systems, Proposed Symbology for. (Committee Report)....................787–93; disc. 793
Diode Logic, Back-Transient. Wolff.............4–9
Diode Logic Gates, Current-Operated. Reinecke..762–72
Diodes, An Investigation of the Dynamic Switching Properties of 4-Layer. McDuffie, Chadwell..50–3
Dissipation Limits in Some Bistable Transistor Pulse Circuits, Fundamental Considerations of Power. Raillard....................................53–5
Distortion at Audio Frequencies with 1-Microsecond Precision, A Test Set for Measurement of Envelope Delay. Codd.................... 241–5
Distortion on Data Transmission, The Effect of Delay. Mertz....................................228–32
Distortion Simulation Set, A Delay. Dunbar....183–5
Distributor, An Electronic Teleprinter. Biswas, Rakshit.. 620–4
Drive System and of Vibrator Drive Systems, An Analysis of a Polarized Electromagnetic. Nunn..346–52
Drum Design, Magnetic Memory. Carne......749–56
Dual-Channel Reactor Protection System for Nuclear Power Plants, A. Bartu....................358–62
Dynamic Electronic Speed-Torque Indicator, Portable. Weed....................................388–92
Dynamic Optimization and Control of a Stirred-Tank Chemical Reactor. Gould, Kipiniak..............................734–43; disc. 743
Dynamic Optimizing Control of the Continuous Process, An Approach to. Sandelien................291–9

E

E6 Telephone Repeater Background, Description, and Theory of Operation, The. Doyle, Talley, Hochgraf, Pedersen, Smethurst....................525–42
Eddy-Current Brake for a Sodium-Cooled Nuclear Power Reactor, Design of an. Baker.......330–3
Education for Expanding Horizons in Electric Power. Brown....................149–53; disc. 153
Educational Television System, Redstone Arsenal Closed-Circuit. Argo, Howell............185–90
Effect of Delay Distortion on Data Transmission, The. Mertz....................................228–32
Effect of Source and Sink Thermal Resistance on Thermoelectric Generator Performance, The. Gray....................................15–19
Effects of Nuclear Radiation on the Electrical Strength of Air, The. Duncan, Fraser, Valachovic...19–26
Electric Analog of a Magnetic Circuit Established by the Basic Laws Defining Magnetostatics, The Existence of an. Koenig....................................400–03
Electric Machine, Instabilities of Push-Pull Magnetic Amplifiers Feeding the Field of an. Storm...33–6
Electric Power, Education for Expanding Horizons in. Brown....................149–53; disc. 153
Electric Power System, Development and Application of Time-Division Multiplexing Telemetering Equipment for Data Transmission on an. Almon, Donelson....................................155–9
(Electrical Insulation) Pigmented Polyester Magnet Wire for High Temperature Operation. Sattler, Scala....................................503–05
(Electrical Insulation) The Effects of Nuclear Radiation on the Electrical Strength of Air. Duncan, Fraser, Valachovic....................................19–26
(Electrical Insulation) The Measurement and Influence of Surface Charges on the Inception Voltage Between Dielectric and Metal Dielectric Surfaces. Reynolds....................................310–14
Electrical Precipitation, A Theory for Space-Charge-Limited Currents with Application to. Cooperman....................................47–50
(Electrical Techniques in Medicine & Biology) Experiments with Air Ions on the Trachea. Beckett, Krueger....................................695–9
Electromagnetic Drive System and of Vibrator Drive Systems, An Analysis of a Polarized. Nunn.346–52
Electromagnets, A Method for the Design of Holding. Ludwig....................................300–10
Electron Tube or a Semiconductor Device Be Used?, Shall an. Scheneman, Waldorf............264–8
Electronic Signalling System, A Military 4-Wire Field Telephone Incorporating an. Faherty, Howell..177–83
Electronic Speed-Torque Indicator, Portable Dynamic. Weed....................................388–92

Electronic Telephone Switchboard, A 240-Line Fully. Goudet......232–41
Electronic Teleprinter Distributor, An. Biswas, Rakshit......620–4
Electronic Transformer Design by Digital Computers. Deise, Etchison, Lee......314–22; disc. 322
(Electronics) A New Ceramic Triode for VHF Applications. Campbell......404–08
(Electronics) A Theory for Space-Charge-Limited Currents with Application to Electrical Precipitation. Cooperman......47–50
(Electronics) Back-Transient Diode Logic. Wolff..4–9
(Electronics) Current-Voltage Relationships in Negative Pulsed Corona. Thomas, Williams......136–9
(Electronics) Experiments on a Series of S-Band Crestatrons. Rowe, Konrad, Krage......828–33
(Electronics) Fundamental Considerations of Power Dissipation Limits in Some Bistable Transistor Pulse Circuits. Raillard......53–5
(Electronics) Locking Range of an Oscillator for Different Nonlinearities. Nag......134–6
(Electronics) Potentials in D-C Corona Fields. Penney, Matick......91–9
(Electronics) Power Relationships and Temperature Dependence in the D-C Corona Field. Thomas, Williams, Suzuki......1–4
(Electronics) Saturable-Reactor Control of Full-Wave and Biphase Rectifiers. Thomas, Drenning..36–9
(Electronics) Sparkover as Influenced by Surface Conditions in D-C Corona. Penney, Craig......112–18
(Electronics) Specifying a Pulse Transformer for Computer Use. Blessing......44–7
(Electronics) Stable Transistor Wide-Band D-C Amplifiers. Okada......26–33
(Electronics) The Empretron, a Mercury Pool Arc Tube Allowing Operation at Repetition Rates in the Kilocycle Range. Hernqvist, Norman..85–91
(Electronics) Tubes or Transistors: A Realistic Assessment. Moe......81–5
(Electronics) Voltage Transients Due to Arc Extinction. Steiner, Strecker......139–44
Eliminators for Teletypewriters, High-Voltage Static. Jonas, Hanscom......576–8
Empretron, a Mercury Pool Arc Tube Allowing Operation at Repetition Rates in the Kilocycle Range, The. Hernqvist, Norman......85–91
Envelope Delay Distortion at Audio Frequencies with 1-Microsecond Precision, A Test Set for Measurement of. Codd......241–5
Equations and Their Applications, Difference. Pipes......333–9; disc. 339
Equipment Engineering, Automation in Switching. Hacker, Davis......545–52
Equipment in Guidance Systems, A Comparison of Digital Differential Analyzer and General Purpose. Dickinson......706–08
Error-Correcting Codes for Correcting Bursts of Errors. Meggitt......708–11
Establishment of a 450-Mc Vehicular Radio System in the Los Angeles Basin. Applegarth......597–606
Excitations, On the Response of Linear Systems to Periodic. LePage, McFee......746–8; disc. 749
Existence of an Electric Analog of a Magnetic Circuit Established by the Basic Laws Defining Magnetostatics, The. Koenig......400–03
Experimental Reactor Kinetic Studies, Instrumentation for. Gardner......384–8
Experiments on a Series of S-Band Crestatrons. Rowe, Konrad, Krage......828–33
Experiments with Air Ions on the Trachea. Beckett, Krueger......695–9

F

FM and PM Systems, Intermodulation Effects in. Love......245–8
Feeding the Field of an Electric Machine, Instabilities of Push-Pull Magnetic Amplifiers. Storm....33–6
Field Telephone Incorporating an Electronic Signalling System, A Military 4-Wire. Faherty, Howell......177–83
57A Microwave Path Protection System, The. Hesselgrave, Kebby......654–9
Filter as Applied to a Time Division Multiplex System, Synthesis of a Resonant Transfer. May, Stump......615–20
Filters with Real or Imaginary Transfer Functions, The Synthesis of Linear. Gilbert, Otterman....323–30
Flow, Generalized Brillouin. Kent......144–8
Focused Cyclotrons, Sector-. Blosser......814–22
Forms Based on Commutative and Associative Binary Operations, Switching Function Canonical. Calingaert......808–14
450-Mc Vehicular Radio System in the Los Angeles Basin, Establishment of a. Applegarth....597–606
Four-Layer Diodes, An Investigation of the Dynamic Switching Properties of. McDuffie, Chadwell......50–3
Four-Wire Field Telephone Incorporating an Electronic Signalling System, A Military. Faherty, Howell......177–83
Frequencies, A Transistor Automatic Gain Control System for Audio. Shirman......664–9
Frequency Interference Considerations in the TD-2 Radio Relay System, Radio-. Curtis......64–70
Frequency Multiplier Circuits, Symmetrical. Dick......125–34
Frequency Transmission Line Engineering, Radio-. Lewis......634–41
Full-Wave and Biphase Rectifiers, Saturable-Reactor Control of. Thomas, Drenning......36–9
Function Generator, A New. Eggers, Sze......756–62
Functional Analysis Problems, The Analog Solution of Some. Graupe......793–7; disc. 798
Fundamental Considerations of Power Dissipation Limits in Some Bistable Transistor Pulse Circuits. Raillard......53–5
Fusion Power from Magnetically Confined Plasma. Mills......833–6

G

Gain Control System for Audio Frequencies, A Transistor Automatic. Shirman......664–9
Gates, Current-Operated Diode Logic. Reinecke......762–72
Generalized Brillouin Flow. Kent......144–8
Generalized Chemical Processing Model for the Investigation of Computer Control, A. Williams, Otto......458–73
Generalizing Communications Systems, Data Processing as a Tool for. Luebbert......431–6
Generation Using Semiconductors and Magnetic Cores, High-Power Pulse. Lassiter, Johannessen, Spencer......511–17
Generator, A New Function. Eggers, Sze......756–62
Generator Performance, The Effect of Source and Sink Thermal Resistance on Thermoelectric. Gray......15–19
Global Communications, Use of Moon or Satellite Relays for. Yeh......607–15
Guidance Systems, A Comparison of Digital Differential Analyzer and General Purpose Equipment in. Dickinson......706–08

H

Harmonic Modulator, An Analysis of the Operation of the Magnetic Second-. Jalbert......268–72
High-Power Pulse Generation Using Semiconductors and Magnetic Cores. Lassiter, Johannessen, Spencer......511–17
High-Speed Microcurrent Amplifier, A. Dever, Sickles......375–9
High-Speed Wide-Chart Recorder, Design and Performance Characteristics of a. Garrigus...163–70
High-Temperature Operation, Pigmented Polyester Magnet Wire for. Sattler, Scala......503–05
High-Voltage Static Eliminators for Teletypewriters. Jonas, Hanscom......576–8
Hoist Control for A-C Powered Cranes, D-C Magnetic Crane. Myles, Davies, Srnka......207–11
Honeywell H-290, Logical Organization of the. Eachus......715–19
Horizons in Electric Power, Education for Expanding. Brown......149–53; disc. 153
Hybrid Computers for Process Control. Birkel......726–34; disc. 734
Hysteresis Loops of Polarized Polycrystalline Barium Titanate, Quantitative Study of Low-Frequency. Marks, Hanna......799–808

I

Ice and Snow Measurement, Proposed Monocycle-Pulse Very-High-Frequency Radar for Air-Borne. Cook......588–93; disc. 593
Impedance of Semiconductor Devices, Transient Thermal. Diebold, Luft......719–26
Impulse Metering System, A New. Snyder, Booker......159–63
Incremental and Conventional Digital Computers in Air-Borne Real-Time Control, The Relative Merits of. Shackell, Tryon......393–400
(Indicating and Integrating Instruments) A New Impulse Metering System. Snyder, Booker......159–63
(Indicating and Integrating Instruments) Thermal Voltage Converters for Accurate Voltage Measurements to 30 Megacycles Per Second. Hermach, Williams......200–05; disc. 206
Indicator, Portable Dynamic Electronic Speed-Torque. Weed......388–92
Inductors Biased with Permanent Magnets. Ludwig:
  Part I......273–8
  Part II......278–91
(Industrial Control) D-C Magnetic Crane Hoist Control for A-C Powered Cranes. Myles, Davies, Srnka......207–11
Industrial Television for Ticket Reservation Facilities, Dial-Selected. Alinsky......174–7
Information by Orthogonal Time Functions, On the Transmission of. Harmuth......248–55
Ink Recorders, Miniaturized Current Transformers for Multirange Operation of Milliammeter. Geyger......681–9
Instabilities of Push-Pull Magnetic Amplifiers Feeding the Field of an Electric Machine. Storm....33–6
Instrumentation for Experimental Reactor Kinetic Studies. Gardner......384–8
Instruments, Protection for Underwater. Edgerton, Hoadley......689–94
Interference Considerations in the TD-2 Radio Relay System, Radio-Frequency. Curtis......64–70
Interference Reduction, Production Design Considerations in Television Chassis Sweep. Carson..557–60
Intermodulation Effects in FM and PM Systems. Love......245–8
Interstitial Channels for Doubling TD-2 Radio System Capacity. Curtis, Collins, Jamison......646–54
Investigation of the Dynamic Switching Properties of 4-Layer Diodes, An. McDuffie, Chadwell...50–3
Ions on the Trachea, Experiments with Air. Beckett, Krueger......695–9
Iterative Networks, Assignment of Carry Variables in. McCluskey......772–8

K

Kilocycle Range, The Empretron, a Mercury Pool Arc Tube Allowing Operation at Repetition Rates in the. Hernqvist, Norman......85–91
Kinetic Studies, Instrumentation for Experimental Reactor. Gardner......384–8

L

LN 3000 Computer Control System, The. Harple, Lex......408–13
Laws Defining Magnetostatics, The Existence of an Electric Analog of a Magnetic Circuit Established by the Basic. Koenig......400–03
Limit Switch for Indicating and Recording Instruments, A Transistor-Oscillator. Wintermute, Hayter......518–19
Linear Filters with Real or Imaginary Transfer Functions, The Synthesis of. Gilbert, Otterman.323–30
Linear Reactor Simulator, A Multidecade. Schwartzenberg, Finan, Underkoffler, Brey......353–8
Linear Systems to Periodic Excitations, On the Response of. LePage, McFee......746–8; disc. 749
Load, Analysis and Optimization of the Bridge Magnetic Servoamplifier with True Motor, Walsh, Meiksin......413–20
Loads, Power Semiconductor Ratings Under Transient and Intermittent. Gutzwiller, Sylvan......699–705; disc. 705
Locking Range of an Oscillator for Different Nonlinearities. Nag......134–6
Logarithmic Count-Rate and Period Meters, Transients in. Barrow, Maitland......362–9
Logic, Back-Transient Diode. Wolff......4–9
Logic Gates, Current-Operated Diode. Reinecke......762–72
Logical Organization of the Honeywell H-290. Eachus......715–19
Los Angeles Basin, Establishment of a 450-Mc Vehicular Radio System in the. Applegarth......597–606
Low-Frequency Hysteresis Loops of Polarized Polycrystalline Barium Titanate, Quantitative Study of. Marks, Hanna......799–808

M

Magnet Coil Design, Some Basic Concepts for. Gauster......822–8
Magnet Wire for High-Temperature Operation, Pigmented Polyester. Sattler, Scala......503–05
Magnetic Amplifier, Theory of 3-Phase Bridge. Storm, Flarity:
  Part II......451–8
  (Part I—1959)
Magnetic Amplifiers Feeding the Field of an Electric Machine, Instabilities of Push-Pull. Storm...33–6
Magnetic Amplifiers, Rectifier Unblocking from a Generalized Analysis of Self-Saturating. Bourne, Nitzan......99–105
(Magnetic Amplifiers) Static Control for a Mechanically Regulated D-C Supply. Abrams, Brubaker......426–30
(Magnetic Amplifiers) Symmetrical Frequency Multiplier Circuits. Dick......125–34
Magnetic Amplifiers, Time Response Constant and Transfer Function of. Manteuffel......473–82; disc. 482
Magnetic Analog with Spiral Search Coils, The. Moore......256–63; disc. 263
Magnetic Automatic Power-Range Control for an Aircraft Nuclear Reactor. Russell, Hemmenway, Scharf, Sharr......379–84

Magnetic Circuit Established by the Basic Laws Defining Magnetostatics, The Existence of an Electric Analog of a. Koenig........400–03
Magnetic Cores, High-Power Pulse Generation Using Semiconductors and. Lassiter, Johannessen, Spencer........511–17
Magnetic-Coupled Multivibrators, Recent Developments on. Geyger........106–12
Magnetic Crane Hoist Control for A-C Powered Cranes, D-C. Myles, Davies, Srnka........207–11
Magnetic Memory Drum Design. Carne........749–56
Magnetic Modulator, Theory and Operation of a Cross-Field. Boensel........505–10; disc. 510
Magnetic Recorder, A Step-Driven Digital. Hansen........564–5
Magnetic Second-Harmonic Modulator, An Analysis of the Operation of the. Jalbert........268–72
Magnetic Servoamplifier with True Motor Load, Analysis and Optimization of the Bridge. Walsh, Meiksin........413–20
Magnetic-Tape-to-Paper-Tape Converter, A. Ringer, Mintzer........339–46
Magnetically Confined Plasma, Fusion Power from. Mills........833–6
Magnetization in Tape-Wound Cores. Barker........482–501; disc. 501
Magnetostatics, The Existence of an Electric Analog of a Magnetic Circuit Established by the Basic Laws Defining. Koenig........400–03
Magnets, Inductor Biased with Permanent. Ludwig:
  Part I........273–8
  Part II........278–91
Matrix Logic VI, Universal Modulus. Schubert........778–81
Measurement and Influence of Surface Charges on the Inception Voltage Between Dielectric and Metal Dielectric Surfaces, The. Reynolds........310–14
Measurement, Telephone Influence Factor (TIF) and Its. Ball, Poarch........659–64
Mechanically Regulated D-C Supply, Static Control for a. Abrams, Brubaker........426–30
Megacycle Telephone Service to North Carolina's Ocracoke Island, Providing 150-. McMillan........542–5
Megacycles Per Second, Thermal Voltage Converters for Accurate Voltage Measurements to 30. Hermach, Williams........200–05; disc. 205
Memory Drum Design, Magnetic. Carne........749–56
Memory Translator with Shared Access, A Wired-. Pace, Ostendorf........216–20
Mercury Pool Arc Tube Allowing Operation at Repetition Rates in the Kilocycle Range, The Empretron. Hernqvist, Norman........85–91
Metered Transatlantic Switching System, A Character-. Coggeshall, Holcomb........56–54; disc. 64
Metering System, A New Impulse. Snyder, Booker........159–63
Meters, Transients in Logarithmic Count-Rate and Period. Barrow, Maitland........362–9
Method for the Design of Holding Electromagnets, A. Ludwig........300–10
Microcurrent Amplifier, A High-Speed. Dever, Sickles........375–9
Microsecond Precision, A Test Set for Measurement of Envelope Delay Distortion at Audio Frequencies with One-. Codd........241–5
Microwave Path Protection System, The 57A. Hesselgrave, Kebby........654–9
Military 4-Wire Field Telephone Incorporating an Electronic Signalling System, A. Faherty, Howell........177–83
Milliammeter Ink Recorders, Miniaturized Current Transformers for Multirange Operation of. Geyger........681–9
Miniaturized Current Transformers for Multirange Operation of Milliammeter Ink Recorders. Geyger........681–9
Mobile Radio System, Centralized Testing of a Very-High-Frequency. MacDonald........641–5
Mobile Telephone Service, A New Control Terminal for. Getgen, McGibbon........583–8
Modulator, An Analysis of the Operation of the Magnetic Second-. Jalbert........268–72
Modulator, Theory and Operation of a Cross-Field Magnetic. Boensel........505–10; disc. 510
Modulus, Universal—Matrix Logic VI. Schubert........778–81
Monocycle-Pulse Very-High-Frequency Radar for Air-Borne Ice and Snow Measurement, Proposed. Cook........588–93; disc. 593
Moon or Satellite Relays for Global Communications, Use of. Yeh........607–15
Moon Reflection Studies with Bistatic Radar at 3,000 Mc. Straiton, Tolbert........436–9; disc. 439
Motor Load, Analysis and Optimization of the Bridge Magnetic Servoamplifier with True. Walsh, Meiksin........413–20
Multidecade Linear Reactor Simulator, A. Schwartzenberg, Finan, Underkoffler, Brey........353–8
Multiplex System, Synthesis of a Reasonant Transfer Filter as Applied to a Time Division. May, Stump........615–20
Multiplexing Telemetering Equipment for Data Transmission on an Electric Power System, Development and Application of Time-Division. Almon, Donelson........155–9
Multiplier Circuits, Symmetrical Frequency. Dick........125–34
Multirange Operation of Milliammeter Ink Recorders, Miniaturized Current Transformers for. Geyger........681–9
Multivibrators, Recent Developments on Magnetic-Coupled. Geyger........106–12

N

N Carrier Systems, Two-Wire Operation of. Dean, Karkar........594–7
Nation-Wide Telephone Numbering Plan, The New. Myers........673–5
Negative Pulsed Corona, Current-Voltage Relationships in. Thomas, Williams........136–9
Network Switching Facilities, New Television. Bachelet, Collins, Taylor........625–34
Networks, Assignment of Carry Variables in Iterative. McCluskey........772–8
New Impulse Metering System, A. Snyder, Booker........159–63
Nondestructive Breakdown Phenomenon in Selenium Rectifiers, A. English, Tobin........9–14
Nonlinear Systems, Analysis of Transcendental. Wolf........449–51
Nonlinearities, Locking Range of an Oscillator for Different. Nag........134–67
North Carolina's Ocracoke Island, Providing 150 Megacycle Telephone Service to. McMillan........542–5
Nuclear Power Plants, A Dual-Channel Reactor Protection System for. Bartu........358–62
Nuclear Power Reactor, Design of an Eddy-Current Brake for a Sodium-Cooled. Baker........330–3
Nuclear Radiation on the Electrical Strength of Air, The Effects of. Duncan, Fraser, Valachovic........19–26
Nuclear Reactor, Magnetic Automatic Power-Range Control for an Aircraft. Russell, Hemmenway, Scharf, Sharr........379–84
Nuclear Reactors, A Digital Start-up Control for Air-Cooled. Lehr, Mathis........369–75
(Nucleonic and Radiation Instrumentation) A High-Speed Microcurrent Amplifier. Dever, Sickles........375–9
(Nucleonic and Radiation Instrumentation) A Multidecade Linear Reactor Simulator. Schwartzenberg, Finan, Underkoffler, Brey........353–8
(Nucleonic and Radiation Instrumentation) Instrumentation for Experimental Reactor Kinetic Studies. Gardner........384–8
(Nucleonic and Radiation Instrumentation) Transients in Logarithmic Count-Rate and Period Meters. Barrow, Maitland........362–9
(Nucleonics) Fusion Power from Magnetically Confined Plasma. Mills........833–6
(Nucleonics) Sector-Focused Cyclotrons. Blosser........814–22
(Nucleonics) Some Basic Concepts for Magnet Coil Design. Gauster........822–8
Numbering Plan, The New Nation-Wide Telephone. Myers........673–5

O

Ocracoke Island, Providing 150-Megacycle Telephone Service to North Carolina's. McMillan........542–5
Omnidirectional Circular Antenna Array Excited Parasitically by a Central Driven Element, An. Neff, Tillman........190–2
Optimization and Control of a Stirred-Tank Chemical Reactor, Dynamic. Gould, Kipiniak........734–43; disc. 743
Optimization of the Bridge Magnetic Servoamplifier with True Motor Load, Analysis and. Walsh, Meiksin........413–20
Optimizing Control of the Continuous Process, An Approach to Dynamic. Sandelien........291–9
Optimum Design Considerations for Radio Relays Utilizing the Tropospheric Scatter Mode of Propagation. Parry........71–80
Orthogonal Time Functions, On the Transmission of Information by. Harmuth........248–55
Orthogonal Time Functions, Radio Communication with. Harmuth........221–8
Oscillator for Different Nonlinearities, Locking Range of an. Nag........134–6
Oscillator Limit Switch for Indicating and Recording Instruments, A Transistor-. Wintermute, Hayter........518–19

P

PM Systems, Intermodulation Effects in FM and. Love........245–8
Panel Wiring, Automation of Computer. Altman, DeCampo, Warburton........118–25
Paper-Tape Converter, A Magnetic-Tape-to-. Ringer, Mintzer........339–46
Parasitically by a Central Driven Element, An Omnidirectional Circular Antenna Array Excited. Neff, Tillman........190–2
Path Protection System, The 57A Microwave. Hesselgrave, Kebby........654–9
Performance Characteristics of a High-Speed Wide-Chart Recorder, Design and. Garrigus........163–70
Period Meters, Transients in Logarithmic Count-Rate and. Barrow, Maitland........362–9
Periodic Excitations, On the Response of Linear Systems to. LePage, McFee........746–8; disc. 749
Permanent Magnets, Inductors Biased with. Ludwig:
  Part I........273–8
  Part II........278–91
Phenomenon in Selenium Rectifiers, A Nondestructive Breakdown. English, Tobin........9–14
Pigmented Polyester Magnet Wire for High-Temperature Operation. Sattler, Scala........503–05
Plan, The New Nation-Wide Telephone Numbering. Myers........673–5
Plasma, Fusion Power from Magnetically Confined. Mills........833–6
Polarized Electromagnetic Drive System and of Vibrator Drive Systems, An Analysis of a. Nunn........346–52
Polarized Polycrystalline Barium Titanate, Quantitative Study of Low-Frequency Hysteresis Loops of. Marks, Hanna........799–808
Pole and Zero Locations for Nondecreasing Step Responses, The Properties of. Zemanian........421–6
Polycrystalline Barium Titanate, Quantitative Study of Low-Frequency Hysteresis Loops of Polarized. Marks, Hanna........799–808
Polyester Magnet Wire for High-Temperature Operation, Pigmented. Sattler, Scala........503–05
Pool Arc Tube Allowing Operation at Repetition Rates in the Kilocycle Range, The Empretron, a Mercury. Hernqvist, Norman........85–91
Portable Dynamic Electronic Speed-Torque Indicator. Weed........388–92
Potentials in D-C Corona Fields. Penney, Matick........91–9
Power Dissipation Limits in Some Bistable Transistor Pulse Circuits, Fundamental Considerations of. Raillard........53–5
Power, Education for Expanding Horizons in Electric. Brown........149–53; disc. 153
Power from Magnetically Confined Plasma, Fusion. Mills........833–6
Power-Line Transient Recorder, A. Hoshall........170–3
Power Plants, A Dual-Channel Reactor Protection System for Nuclear. Bartu........358–62
Power-Range Control for an Aircraft Nuclear Reactor, Magnetic Automatic. Russell, Hemmenway, Sharf, Sharr........379–84
Power Relationships and Temperature Dependence in the D-C Corona Field. Thomas, Williams, Suzuki........1–4
Power Semiconductor Ratings Under Transient and Intermittent Loads. Gutzwiller, Sylvan........699–705; disc. 705
Power System, Development and Application of Time-Division Multiplexing Telemetering Equipment for Data Transmission on an Electric. Almon, Donelson........155–9
Precipitation, A Theory for Space-Charge-Limited Currents with Application to Electrical. Cooperman........47–50
Precision, Continuous Voltage Reference for Industrial Recorders, A. Robinson........197–200
Problems, The Analog Solution of Some Functional Analysis. Graupe........793–7; disc. 798
Process Control, Hybrid Computers for. Birkel........726–34; disc. 734
Processing as a Tool for Generalizing Communications Systems, Data. Luebbert........431–6
Processing Model for the Investigation of Computer Control, A Generalized Chemical. Williams, Otto........458–73
Production Design Considerations in Television Chassis Sweep Interference Reduction. Carson........557–60
Propagation, Optimum Design Considerations for Radio Relays Utilizing the Tropospheric Scatter Mode of. Parry........71–80
Properties of Pole and Zero Locations for Nondecreasing Step Responses, The. Zemanian........421–6
Proposal for New Concepts in Digital Data Processing, The Combi-System—A. Schwab........193–7
Proposed Monocycle-Pulse Very-High-Frequency Radar for Air-Borne Ice and Snow Measurement. Cook........588–93; disc. 593
Proposed Symbology for Digital Systems. (Committee Report)........787–93; disc. 793
Protection for Underwater Instruments. Edgerton, Hoadley........689–94
Protection System for Nuclear Power Plants, A Dual-Channel Reactor. Bartu........358–62
Protection System, The 57A Microwave Path. Hesselgrave, Kebby........654–9

Providing 150-Megacycle Telephone Service to North Carolina's Ocracoke Island. McMillan.....542–5
Pulse Circuits, Fundamental Considerations of Power Dissipation Limits in Some Bistable Transistor. Raillard.....53–5
Pulse Generation Using Semiconductors and Magnetic Cores, High-Power. Lassiter, Johannessen, Spencer.....511–17
Pulse Transformer for Computer Use, Specifying a. Blessing.....44–7
Pulse Very-High-Frequency Radar for Air-Borne Ice and Snow Measurement, Proposed Monocycle-. Cook.....588–93; disc. 593
Pulsed Corona, Current-Voltage Relationships in Negative. Thomas, Williams.....136–9
Push-Pull Magnetic Amplifiers Feeding the Field of an Electric Machine, Instabilities of. Storm....33–6

Q

Quantitative Study of Low-Frequency Hysteresis Loops of Polarized Polycrystalline Barium Titanate. Marks, Hanna.....799–808

R

REA-1 Transposition System, The. Chapman, Buscemi, Flanigan.....440–8
Radar at 3,000 Mc, Moon Reflection Studies with Bistatic. Straiton, Tolbert.......436; disc. 439
Radar for Air-Borne Ice and Snow Measurement, Proposed Monocycle-Pulse Very-High-Frequency. Cook.....588–93; disc. 593
Radiation on the Electrical Strength of Air, The Effects of Nuclear. Duncan, Fraser, Valachovic....19–26
(Radio Communication Systems) A New Control Terminal for Mobile Telephone Service. Getgen, McGibbon.....583–8
(Radio Communication Systems) Intermodulation Effects in FM and PM Systems. Love......245–8
(Radio Communication Systems) Moon Reflection Studies with Bistatic Radar at 3,000 Mc. Straiton, Tolbert.....436–9; disc. 439
(Radio Communication Systems) Providing 150-Megacycle Telephone Service to North Carolina's Ocracoke Island. McMillan.....542–5
(Radio Communication Systems) The 57A Microwave Path Protection System. Hesselgrave, Kebby.....654–9
(Radio Communication Systems) Use of Moon or Satellite Relays for Global Communications. Yeh.....507–15
Radio Communication with Orthogonal Time Functions. Harmuth.....221–8
Radio-Frequency Interference Considerations in the TD-2 Radio Relay System. Curtis.....64–70
Radio-Frequency Transmission Line Engineering. Lewis.....634–41
Radio Relays Utilizing the Tropospheric Scatter Mode of Propagation, Optimum Design Considerations for. Parry.....71–80
Radio System Capacity, Interstitial Channels for Doubling TD-2. Curtis, Collins, Jamison..646–54
Radio System, Centralized Testing of a Very-High-Frequency Mobile. MacDonald.....641–5
Radio System in the Los Angeles Basin, Establishment of a 450-Mc Vehicular. Applegarth......597–606
Railroad, Wire Line and Carrier Communications on the Union Pacific. Jett.....520–5
Ratings Under Transient and Intermittent Loads, Power Semiconductor. Gutzwiller, Sylvan.....699–705; disc. 705
Reactor, Design of an Eddy-Current Brake for a Sodium-Cooled Nuclear Power. Baker.....330–3
Reactor Kinetic Studies, Instrumentation for Experimental. Gardner.....384–8
Reactor Protection System for Nuclear Power Plants, A Dual-Channel. Bartu.....358–62
Reactor Simulator, A Multidecade Linear. Schwartzenberg, Finan, Underkoffler, Brey.....353–8
Recent Developments on Magnetic-Coupled Multivibrators. Geyger.....106–12
Recorder, A Power-Line Transient. Hoshall....170–3
Recorder, A Step-Driven Digital Magnetic. Hansen.....564–5
Recorder, Design and Performance Characteristics of a High-Speed Wide-Chart. Garrigus.......163–70
Recorder with Emphasis on Traffic Engineering, Reflections on the Traffic Usage. Bauerfeld.....566–70
Recorders, A Precision, Continuous Voltage Reference for Industrial. Robinson.....197–200
Recorders, Miniaturized Current Transformers for Multirange Operation of Milliammeter Ink. Geyger.....681–9
(Recording and Controlling Instrumentation) The Combi-System—A Proposal for New Concepts in Digital Data Processing. Schwab.....193–7
Recording Instruments, A Transistor-Oscillator Limit Switch for Indicating and. Wintermute, Hayter.....518–19
Rectifier Unblocking from a Generalized Analysis of Self-Saturating Magnetic Amplifiers. Bourne, Nitzan.....99–105
Rectifiers, Saturable-Reactor Control of Full-Wave and Biphase. Thomas, Drenning.....36–9
Redstone Arsenal Closed-Circuit Educational Television System. Argo, Howell.....185–90
Reflection Studies with Bistatic Radar at 3,000 Mc, Moon. Straiton, Tolbert.......436–9; disc. 439
Reflections on the Traffic Usage Recorder with Emphasis on Traffic Engineering. Bauerfeld.......566–70
Regulated D-C Supply, Static Control for a Mechanically. Abrams, Brubaker.....426–30
Relative Merits of Incremental and Conventional Digital Computers in Air-Borne Real-Time Control, The. Shackell, Tryon.........393–400
Relay System, Radio-Frequency Interference Considerations in the TD-2 Radio. Curtis.....64–70
Relays for Global Communications, Use of Moon or Satellite. Yeh.....607–15
Relays Utilizing the Tropospheric Scatter Mode of Propagation, Optimum Design Considerations for Radio. Parry.....71–80
Repeater, A Transistorized Cable Carrier. Griffith.....675–80
Repeater Background, Description, and Theory of Operation, The E6 Telephone. Doyle, Talley, Hochgraf, Pedersen, Smethurst.....525–42
Repetition Rates in the Kilocycle Range, The Empretron, a Mercury Pool Arc Tube Allowing Operation at. Hernqvist, Norman.....85–91
Reservation Facilities, Dial-Selected Industrial Television for Ticket. Alinsky.....174–7
Reservations, Teletypewriter Selective Calling Applied to Automatic Airlines. Bratschi, Pennington, Whitman.....579–83
Resonant Transfer Filter as Applied to a Time Division Multiplex System, Synthesis of a. May, Stump.....615–20
Response of Linear Systems to Periodic Excitations, On the. LePage, McFee..........746–8; disc. 749
Responses, The Properties of Pole and Zero Locations for Nondecreasing Step. Zemanian........421–6

S

S-Band Crestatrons, Experiments on a Series of. Rowe, Konrad, Krage.....828–33
Satellite Relays for Global Communications, Use of Moon or. Yeh.....607–15
Saturable-Reactor Control of Full-Wave and Biphase Rectifiers. Thomas, Drenning.....36–9
Scatter Mode of Propagation, Optimum Design Considerations for Radio Relays Utilizing the Tropospheric. Parry.....71–80
Second-Harmonic Modulator, An Analysis of the Operation of the Magnetic. Jalbert......268–72
Sector-Focused Cyclotrons. Blosser.....814–22
Selective Calling Applied to Automatic Airlines Reservations, Teletypewriter. Bratschi, Pennington, Whitman.....579–83
Selenium Rectifiers, A Nondestructive Breakdown Phenomenon in. English, Tobin.....9–14
Self-Saturating Magnetic Amplifiers, Rectifier Unblocking from a Generalized Analysis of. Bourne, Nitzan.....99–105
Semiconductor Device Be Used?, Shall an Electron Tube or a. Scheneman, Waldorf.....264–8
Semiconductor Devices, Transient Thermal Impedance of. Diebold, Luft.....719–26
(Semiconductor Metallic Rectifiers) An Investigation of the Dynamic Switching Properties of 4-Layer Diodes. McDuffie, Chadwell.....50–3
Semiconductor Ratings Under Transient and Intermittent Loads, Power. Gutzwiller, Sylvan.....699–705; disc. 705
Semiconductors and Magnetic Cores, High-Power Pulse Generation Using. Lassiter, Johannessen, Spencer.....511–17
Series of S-Band Crestatrons, Experiments on a. Rowe, Konrad, Krage.....828–33
Servoamplifier with True Motor Load, Analysis and Optimization of the Bridge Magnetic. Walsh, Meiksin.....413–20
Set Using Swept-Band Techniques, A Voice-Band Transmission-Measuring. Lent.....570–6
Shall an Electron Tube or a Semiconductor Device Be Used? Scheneman, Waldorf.....264–8
Signalling System, A Military 4-Wire Field Telephone Incorporating an Electronic. Faherty, Howell.....177–83
Simulation, A 2-Channel Data Link for Combined Analog-Digital. Greenstein.....40–4
Simulation Set, A Delay Distortion. Dunbar....183–5
Simulator, A Multidecade Linear Reactor. Schwartzenberg, Finan, Underkoffler, Brey.....353–8
Sink Thermal Resistance on Thermoelectric Generator Performance, The Effect of Source and. Gray.....15–19
Snow Measurement, Proposed Monocycle-Pulse Very-High-Frequency Radar for Air-Borne Ice and. Cook.....588–93; disc. 593
Sodium-Cooled Nuclear Power Reactor, Design of an Eddy-Current Brake for a. Baker.....330–3
(Solid State Devices) A Nondestructive Breakdown Phenomenon in Selenium Rectifiers. English, Tobin.....9–14
(Solid State Devices) The Effect of Source and Sink Thermal Resistance on Thermoelectric Generator Performance. Gray.....15–19
Solution of Some Functional Analysis Problems, The Analog. Graupe.....793–7; disc. 798
Space-Charge-Limited Currents with Application to Electrical Precipitation, A Theory for. Cooperman.....47–50
Sparkover as Influenced by Surface Conditions in D-C Corona. Penney, Craig.....112–18
Specifying a Pulse Transformer for Computer Use. Blessing.....44–7
Speed-Torque Indicator, Portable Dynamic Electronic. Weed.....388–92
Spiral Search Coils, The Magnetic Analog with. Moore.....256–63
Stable Transistor Wide-Band D-C Amplifiers. Okada.....26–33
Start-up Control for Air-Cooled Nuclear Reactors, A Digital. Lehr, Mathis.....369–75
Static Control for a Mechanically Regulated D-C Supply. Abrams, Brubaker.....426–30
Static Eliminators for Teletypewriters, High-Voltage. Jonas, Hanscom.....576–8
Step-Driven Digital Magnetic Recorder, A. Hansen.....564–5
Step Responses, The Properties of Pole and Zero Locations for Nondecreasing. Zemanian.....421–6
Stirred-Tank Chemical Reactor, Dynamic Optimization and Control of a. Gould, Kipiniak.....734–43; disc. 743
Studies, Instrumentation for Experimental Reactor Kinetic. Gardner.....384–8
Studies with Bistatic Radar at 3,000 Mc, Moon Reflection. Straiton. Tolbert......436–9; disc. 439
Study of Low-Frequency Hysteresis Loops of Polarized Polycrystalline Barium Titanate, Quantitative. Marks, Hanna.....799–808
Subset, Datacom Unit—An Experimental Data Transmission. Stoffels.....669–73
Surface Conditions in D-C Corona, Sparkover as Influenced by. Penney, Craig.....112–18
Swept-Band Techniques, A Voice-Band Transmission-Measuring Set Using. Lent.....570–6
Switch for Indicating and Recording Instruments, A Transistor-Oscillator Limit. Wintermute, Hayter.....518–19
Switchboard, A 240-Line Fully Electronic Telephone. Goudet.....232–41
Switching Centers, Voice Immunity of Tone Operated. Guenther, Shirman.....553–7
Switching Equipment Engineering, Automation in. Hacker, Davis.....545–52
Switching Equipment Engineering, Automation of Telephone. Blumer, Reifenstuhl, List......560–4
Switching Facilities, New Television Network. Bachelet, Collins, Taylor.....625–34
Switching Function Canonical Forms Based on Commutative and Associative Binary Operations. Calingaert.....808–14
Switching Properties of 4-Layer Diodes, An Investigation of the Dynamic. McDuffie, Chadwell.......50–3
Switching System, A Character-Metered Transatlantic. Coggeshall, Holcomb............56–64; disc. 64
Symbology for Digital Systems, Proposed. (Committee Report).....787–93; disc. 793
Symmetrical Frequency Multiplier Circuits. Dick.....125–34
Synthesis of a Resonant Transfer Filter as Applied to a Time Division Multiplex System. May, Stump.....615–20
Synthesis of Linear Filters with Real or Imaginary Transfer Functions, The. Gilbert, Otterman.....323–30
System, The LN 3000 Computer Control. Harple, Lex.....408–13
System, The REA-1 Transposition. Chapman, Buscemi, Flanigan.....440–8

T

TD-2 Radio Relay System, Radio-Frequency Interference Considerations in the. Curtis.......64–70
TD-2 Radio System Capacity, Interstitial Channels for Doubling. Curtis, Collins, Jamison.......646–54
Tank Chemical Reactor, Dynamic Optimization and Control of a Stirred-. Gould, Kipiniak.....734–43; disc. 743
Tape-to-Paper-Tape Converter, A Magnetic-. Ringer, Mintzer.....339–46
Tape-Wound Cores, Magnetization in. Barker.....482–501; disc. 501
Teaching the Uses of Digital Computers at the Technical University of Denmark. Franksen.....781–7
(Telegraph Systems) A Character-Metered Transatlantic Switching System. Coggeshall, Holcomb.....56–64; disc. 64

(Telegraph Systems) A Step-Driven Digital Magnetic Recorder. Hansen.........564–5
(Telegraph Systems) An Electronic Teleprinter Distributor. Biswas, Rakshit.........620–4
(Telegraph Systems) High-Voltage Static Eliminators for Teletypewriters. Jonas, Hanscom.........576–8
Telemetering Equipment for Data Transmission on an Electric Power System, Development and Application of Time-Division Multiplexing. Almon, Donelson.........155–9
Telephone Incorporating an Electronic Signalling System, A Military 4-Wire Field. Faherty, Howell.........177–83
Telephone Influence Factor (TIF) and Its Measurement. Ball, Poarch.........659–64
Telephone Numbering Plan, The New Nation-Wide. Myers.........673–5
Telephone Repeater Background, Description, and Theory of Operation, The E6. Doyle, Talley, Hochgraf, Pedersen, Smethurst.........525–42
Telephone Service, A New Control Terminal for Mobile Getgen, McGibbon.........583–8
Telephone Service to North Carolina's Ocracoke Island, Providing 150-Megacycle. McMillan.........542–5
Telephone Switchboard, A 240-Line Fully Electronic. Goudet.........232–41
Telephone Switching Equipment Engineering, Automation of. Blumer, Reifenstuhl, List.........560–4
Teleprinter Distributor, An Electronic. Biswas, Rakshit.........620–4
Teletypewriter Selective Calling Applied to Automatic Airlines Reservations. Bratschi, Pennington, Whitman.........579–83
Teletypewriters, High-Voltage Static Eliminators for. Jonas, Hanscom.........576–8
Television Chassis Sweep Interference Reduction, Production Design Considerations in. Carson.........557–60
Television for Ticket Reservation Facilities, Dial-Selected Industrial. Alinsky.........174–7
Television Network Switching Facilities, New. Bachelet, Collins, Taylor.........625–34
Television System, Redstone Arsenal Closed-Circuit Educational. Argo, Howell.........185–90
Temperature Dependence in the D-C Corona Field, Power Relationships and. Thomas, Williams, Suzuki.........1–4
Terminal for Mobile Telephone Service, A New Control. Getgen, McGibbon.........583–8
Test Set for Measurement of Envelope Delay Distortion at Audio Frequencies with 1-Microsecond Precision, A. Codd.........241–5
Testing of a Very-High-Frequency Mobile Radio System, Centralized. MacDonald.........641–5
Theory and Operation of a Cross-Field Magnetic Modulator. Boensel.........505–10; disc. 510
Theory for Space-Charge-Limited Currents with Application to Electrical Precipitation, A. Cooperman.........47–50
Theory of 3-Phase Bridge Magnetic Amplifier. Storm, Flairty:
  Part II.........451–8
  (Part I—1959)
Thermal Resistance on Thermoelectric Generator Performance, The Effect of Source and Sink. Gray.........15–19
Thermal Voltage Converters for Accurate Voltage Measurements to 30 Megacycles Per Second. Hermach, Williams.........200–05; disc. 205
Thermoelectric Generator Performance, The Effect of Source and Sink Thermal Resistance on. Gray.........15–19
30 Megacycles Per Second, Thermal Voltage Converters for Accurate Voltage Measurements to. Hermach, Williams.........200–05; disc. 205
Three-Phase Bridge Magnetic Amplifier, Theory of. Storm, Flairty:
  Part II.........451–8
  (Part I—1959)
3,000 Mc, Moon Reflection Studies with Bistatic Radar at. Straiton, Tolbert.........436–9; disc. 439
Ticket Reservation Facilities, Dial-Selected Industrial Television for. Alinksy.........174–7
Time Division Multiplex System, Synthesis of a Resonant Transfer Filter as Applied to a. May, Stump.........615–20
Time-Division Multiplexing Telemetering Equipment for Data Transmission on an Electric Power System, Development and Application of. Almon, Donelson.........155–9
Time Functions, On the Transmission of Information by Orthogonal. Harmuth.........248–55
Time Functions, Radio Communication with Orthogonal. Harmuth.........221–8
Time Response Constant and Transfer Function of Magnetic Amplifiers. Manteuffel.........473–82; disc. 482
Titanate, Quantitative Study of Low-Frequency Hysteresis Loops of Polarized Polycrystalline Barium. Marks, Hanna.........799–808
Tone Operated Switching Centers, Voice Immunity of. Guenther, Shirman.........553–7
Trachea, Experiments with Air Ions on the. Beckett, Krueger.........695–9
Traffic Usage Recorder with Emphasis on Traffic Engineering. Reflections on the. Bauerfeld.........566–70
Transatlantic Switching System, A Character-Metered. Coggeshall, Holcomb.........56–64; disc. 64
Transcendental Nonlinear Systems, Analysis of. Wolf.........449–51
Transfer Filter as Applied to a Time Division Multiplex System, Synthesis of a Resonant. May, Stump.........615–20
Transfer Function of Magnetic Amplifiers, Time Response Constant and. Manteuffel.........473–82; disc. 482
Transfer Functions, The Synthesis of Linear Filters with Real or Imaginary. Gilbert, Otterman.........323–30
Transformer Design by Digital Computers, Electronic. Deise, Etchison, Lee.........314–22; disc. 322
Transformer for Computer Use, Specifying a Pulse. Blessing.........44–7
Transformers for Multirange Operation of Milliammeter Ink Recorders, Miniaturized Current. Geyger.........681–9
Transient and Intermittent Loads, Power Semiconductor Ratings Under. Gutzwiller, Sylvan.........699–705; disc. 705
Transient Diode Logic, Back-. Wolff.........4–9
Transient Recorder, A Power-Line. Hoshall.........170–3
Transient Thermal Impedance of Semiconductor Devices. Diebold, Luft.........719–26
Transients Due to Arc Extinction, Voltage. Steiner, Strecker.........139–44
Transients in Logarithmic Count-Rate and Period Meters. Barrow, Maitland.........362–9
Transistor Automatic Gain Control System for Audio Frequencies, A. Shirman.........664–9
Transistor-Oscillator Limit Switch for Indicating and Recording Instruments, A. Wintermute, Hayter.........518–19
Transistor Pulse Circuits, Fundamental Considerations of Power Dissipation Limits in Some Bistable. Raillard.........53–5
Transistor Wide-Band D-C Amplifiers, Stable. Okada.........26–33
Transistorized Cable Carrier Repeater, A. Griffith.........675–80
Transistorized Trunk Carrier System, An All-. Gardner.........212–16
Transistors: A Realistic Assessment, Tubes or. Moe.........81–5
Translator with Shared Access, A Wired-Memory. Pace, Ostendorf.........216–20
Transmission Line Engineering, Radio-Frequency. Lewis.........634–41
Transmission-Measuring Set Using Swept-Band Techniques, A Voice-Band. Lent.........570–6
Transmission of Information by Orthogonal Time Functions, On the. Harmuth.........248–55
Transmission Subset, Datacom Unit—An Experimental Data. Stoffels.........669–73
Triode for VHF Applications, A New Ceramic. Campbell.........404–08
Tropospheric Scatter Mode of Propagation, Optimum Design Considerations for Radio Relays Utilizing the. Parry.........71–80
Trunk Carrier System, An All-Transistorized. Gardner.........212–16
Tube Allowing Operation at Repetition Rates in the Kilocycle Range, The Empretron, a Mercury Pool Arc. Hernqvist, Norman.........85–91
Tubes or Transistors: A Realistic Assessment. Moe.........81–5
Two-Channel Data Link for Combined Analog-Digital Simulation, A. Greenstein.........40–4
240-Line Fully Electronic Telephone Switchboard, A. Goudet.........232–41
Two-Wire Operation of N Carrier Systems. Dean, Karkar.........594–7

U

Unblocking from a Generalized Analysis of Self-Saturating Magnetic Amplifiers, Rectifier. Bourne, Nitzan.........99–105
Underwater Instruments, Protection for. Edgerton, Hoadley.........689–94
Union Pacific Railroad, Wire Line and Carrier Communications on the. Jett.........520–5
Universal Modulus—Metrix Logic VI. Schubert.........778–81
Usage Recorder with Emphasis on Traffic Engineering, Reflections on the Traffic. Bauerfeld.........566–70

V

VHF Applications, A New Ceramic Triode for. Campbell.........404–08
Variables in Iterative Networks, Assignment of Carry. McCluskey.........772–8
Vehicular Radio System in the Los Angeles Basin, Establishment of a 450-Mc. Applegarth.........597–606
Very-High-Frequency Mobile Radio System, Centralized Testing of a. MacDonald.........641–5
Vibrator Drive Systems, An Analysis of a Polarized Electromagnetic Drive System and of. Nunn.........364–52
Voice-Band Transmission-Measuring Set Using Swept-Band Techniques, A. Lent.........570–6
Voice Immunity of Tone Operated Switching Centers. Guenther, Shirman.........553–7
Voltage Between Dielectric and Metal Dielectric Surfaces, The Measurement and Influence of Surface Charges on the Inception. Reynolds.........310–14
Voltage Converters for Accurate Voltage Measurements to 30 Megacycles Per Second, Thermal. Hermach, Williams.........200–05; disc. 205
Voltage Reference for Industrial Recorders, A Precision, Continuous. Robinson.........197–200
Voltage Relationships in Negative Pulsed Corona, Current-. Thomas, Williams.........136–9
Voltage Transients Due to Arc Extinction. Steiner, Strecker.........139–44

W

Wide-Band D-C Amplifiers, Stable Transistor. Okada.........26–33
Wide-Chart Recorder, Design and Performance Characteristics of a High-Speed. Garrigus.........163–70
(Wire Communication Systems) A Delay Distortion Simulation Set. Dunbar.........183–5
(Wire Communication Systems) A Military 4-Wire Field Telephone Incorporating an Electronic Signalling System. Faherty, Howell.........177–83
(Wire Communications Systems) A Transistor Automatic Gain Control System for Audio Frequencies. Shirman.........664–9
(Wire Communication Systems) A Transistorized Cable Carrier Repeater. (60–1251).........675–80
(Wire Communication Systems) A Voice-Band Transmission-Measuring Set Using Swept-Band Techniques. Lent.........570–6
(Wire Communication Systems) An All-Transistorized Trunk Carrier System. Gardner.........212–16
(Wire Communication Systems) Telephone Influence Factor (TIF) and Its Measurement. Ball, Poarch.........659–64
(Wire Communication Systems) The E6 Telephone Repeater Background, Description, and Theory of Operation. Doyle, Talley, Hochgraf, Pedersen, Smethurst.........626–42
(Wire Communication Systems) The Effect of Delay Distortion on Data Transmission. Mertz.........228–32
(Wire Communication Systems) The REA-1 Transposition System. Chapman, Buscemi, Flanigan.........440–8
Wire Line and Carrier Communications on the Union Pacific Railroad. Jett.........520–5
Wire Operation of N Carrier Systems, Two-. Dean, Karkar.........594–7
Wired-Memory Translator with Shared Access, A. Pace, Ostendorf.........216–20
Wiring, Automation of Computer Panel. Altman, DeCampo, Warburton.........118–25

Z

Zero Locations for Nondecreasing Stop Responses, The Properties of Pole and. Zemanian.........421–6

2. Author Index

Abrams, H. J.; J. F. Brubaker. Static Control for a Mechanically Regulated D-C Supply. (60–783).........426–30
Alinsky, J. W. Dial-Selected Industrial Television for Ticket Reservation Facilities. (60–260).........174–7
Almon, C. P., Jr.; J. Donelson, Jr. Development and Application of Time-Division Multiplexing Telemetering Equipment for Data Transmission on an Electric Power System. (60–245).........155–9
Altman, G. W.; L. A. DeCampo, C. R. Warburton. Automation of Computer Panel Wiring. (60–128).........118–25
Applegarth, G. Establishment of a 450-Mc Vehicular Radio System in the Los Angeles Basin. (60–995).........597–606
Argo, J. P.; I. N. Howell. Redstone Arsenal Closed-Circuit Educational Television System. (60–59).........185–90

B

Bachelet, A. E.; C. A. Collins, E. R. Taylor. New Television Network Switching Facilities. (60–811).........625–34

Baker, R. S. Design of an Eddy-Current Brake for a Sodium-Cooled Nuclear Power Reactor. (60–150) 330–3

Ball, W. C.; C. K. Poarch. Telephone Influence Factor (TIF) and Its Measurement. (60–1195)...659–64

Barker, R. C. Magnetization in Tape-Wound Cores. (50–873)....482–501; disc. 502

Barrow, Bruce B.; R. Maitland. Transients in Logarithmic Count-Rate and Period Meters. (60–511) 362–9

Bartu, Allen S. A Dual-Channel Reactor Protection System for Nuclear Power Plants. (60–78)........358–62

Bauerfeld, R. A. Reflections on the Traffic Usage Recorder with Emphasis on Traffic Engineering. (60–819)....566–70

Beckett, John C.; A. P. Krueger. Experiments with Air Ions on the Trachea. (60–101)........695–9

Bills, G. W. Amplidyne Circuit Analysis. (60–1006).....711–15

Birkel, George, Jr. Hybrid Computers for Process Control. (60–978)....726–34; disc. 734

Biswas, N. N.; H. Rakshit. An Electronic Teleprinter Distributor. (60–770)....620–4

Blessing, R. R. Specifying a Pulse Transformer for Computer Use. (59–119)....44–7

Blodgett, F. W. Disc....837

Blosser, H. G. Sector-Focused Cyclotrons. (60–1408) 814–22

Blumer, L. W.; G. E. Reifenstuhl, N. E. List. Automation of Telephone Switching Equipment Engineering. (60–817)....560–4

Boensel, D. W. Theory and Operation of a Cross-Field Magnetic Modulator. (60–788)........505–10; disc. 510

Booker, C. A., Jr.; C. J. Snyder. A New Impulse Metering System. (60–64)....159–63

Bourne, H. C.; D. Nitzan. Rectifier Unblocking from a Generalized Analysis of Self-Saturating Magnetic Amplifiers. (60–91)....99–105

Bratschi, R. W.; W. D. Pennington, A. L. Whitman. Teletypewriter Selective Calling Applied to Automatic Airlines Reservations. (60–861).....579–83

Brey, R. N., Jr.; J. W. Schwartzenberg, J. M. Finan, V. S. Underkoffler. A Multidecade Linear Reactor Simulator. (60–516)....353–8

Brown, Gordon S. Education for Expanding Horizons in Electric Power. (60–17)......149–53; disc. 154

Brubaker, J. F.; H. J. Abrams. Static Control for a Mechanically Regulated D-C Supply. (60–783).....426–30

Bullock, W. C. Disc....501

Buscemi, J. V.; A. G. Chapman, J. M. Flanigan. The REA-1 Transposition System. (60–43).....440–8

C

Calingaert, Peter. Switching Function Canonical Forms Based on Commutative and Associative Binary Operations. (60–1199)....808–14

Campbell, J. D. A New Ceramic Triode for VHF Applications. (60–75)....404–08

Carne, E. B. Magnetic Memory Drum Design. (60–1200)....749–56

Carson, G. W. Production Design Considerations in Television Chassis Sweep Interference Reduction. (60–812)....557–60

Chadwell, W. L.; G. E. McDuffie, Jr. An Investigation of the Dynamic Switching Properties of 4-Layer Diodes. (60–19)....50–3

Chapman, A. G.; J. V. Buscemi, J. M. Flanigan. The REA-1 Transposition System. (60–43).....440–8

Clymer, A. B.; T. F. Potts, W. C. Franke, C. R. Walli, G. N. Ornstein, B. J. Miller. Disc....798

Codd, W. Alfred. A Test Set for Measurement of Envelope Delay Distortion at Audio Frequencies with 1-Microsecond Precision. (60–46).....241–5

Coggeshall, I. S.; P. Holcomb, Jr. A Character-Metered Transatlantic Switching System. (58–1143)....56–64

Collins, C. A.; A. E. Bachelet, E. R. Taylor. New Television Network Switching Facilities. (60–811) 625–34

Collins, T. R. D.; H. E. Curtis, B. C. Jamison. Interstitial Channels for Doubling TD-2 Radio System Capacity. (60–1248)....646–54

Cook, John C. Proposed Monocycle-Pulse Very-High Frequency Radar for Air-Borne Ice and Snow Measurement. (60–994)........588–93; disc. 593

Cooley, Austin G. Disc....56

Cooperman, P. A Theory for Space-Charge-Limited Currents with Application to Electrical Precipitation. (60–2)....47–50

Craig, Samuel E.; G. W. Penney. Sparkover as Influenced by Surface Conditions in D-C Corona. (60–103)....112–18

Curtis, H. E. Radio-Frequency Interference Considerations in the TD-2 Radio Relay System. (60–54)....64–70

Curtis, H. E.; T. R. D. Collins, B. C. Jamison. Interstitial Channels for Doubling TD-2 Radio System Capacity. (60–1248)....646–54

D

Davies, M. C.; A. H. Myles, L. J. Srnka. D-C Magnetic Crane Hoist Control for A-C Powered Cranes. (60–30)....207–11

Davis, C. M.; M. E. Hacker. Automation in Switching Equipment Engineering. (60–778)........545–52

Dean, R. W.; E. M. Karkar. Two-Wire Operation of N Carrier Systems. (60–998)....594–7

DeCampo, L. A.; G. W. Altman, C. R. Warburton. Automation of Computer Panel Wiring. (60–128) 118–25

Deise, L. F.; W. Etchison, R. Lee. Electronic Transformer Design by Digital Computers. (59–120).....314–22; disc. 322

Dever, John A.; L. Sickles. A High-Speed Microcurrent Amplifier. (60–513)....375–9

Dick, George W. Symmetrical Frequency Multiplier Circuits. (60–66)....125–34

Dickey, Frank R., Jr. Disc....593

Dickinson, M. M. A Comparison of Digital Differential Analyzer and (General Purpose Equipment in Guidance Systems. 60–122)....706–08

Diebold, E. J.; W. Luft. Transient Thermal Impedance of Semiconductor Devices. (60–68)...719–26

Diebold, E. J.; W. Luft. Disc....705

Donelson, John, Jr.; C. P. Almon, Jr. Development and Application of Time-Division Multiplexing Telemetering Equipment for Data Transmission on an Electric Power System. (60–245).....155–9

Doyle, E. J.; T. J. Talley, L. Hochgraf, L. Pedersen, J. O. Smethurst. The E6 Telephone Repeater Background, Description, and Theory of Operation. (60–814)....525–42

Drenning, J. W.; J. B. Thomas. Saturable-Reactor Control of Full-Wave and Biphase Rectifiers. (59–223)....36–9

Dunbar, Frances C. A Delay Distortion Simulation Set. (60–45)....183–5

Duncan, G. I.; J. C. Fraser, B. Valachovic. The Effects of Nuclear Radiation on the Electrical Strength of Air. (59–134)....19–26

E

Eachus, Joseph J. Logical Organization of the Honeywell H-290. (60–971)....715–19

Edgerton, Harold E.; L. D. Hoadley. Protection for Underwater Instruments. (60–662)........689–94

Eggers, C. W.; T. W. Sze. A New Function Generator. (60–1203)....756–62

English, A. C.; W. H. Tobin. A Nondestructive Breakdown Phenomenon in Selenium Rectifiers. (58–831)....9–14

Estcourt, V. F. Disc....153

Estrin, Gerald; A. B. Rosenstein, A. H. Hadley. Disc. 322

Etchison, Wade; L. F. Deise, R. Lee. Electronic Transformer Design By Digital Computers. (59–120).....314–22; disc. 322

F

Faherty, J. L.; A. S. Howell. A Military 4-Wire Field Telephone Incorporating an Electronic Signalling System. (59–259)....177–83

Finan, J. M.; J. W. Schwartzenberg, V. S. Underkoffler, R. N. Brey, Jr. A Multidecade Linear Reactor Simulator. (60–516)....353–8

Flairty, Clarence W.; H. F. Storm. Theory of 3-Phase Bridge Magnetic Amplifier:
  Part II. (60–872)....451–8
  (Part I—1959)

Flanigan, J. M.; A. G. Chapman, J. V. Buscemi. The REA-1 Transposition System. (60–43).....440–8

Franke, W. C.; A. B. Clymer, T. F. Potts, C. R. Walli, G. N. Ornstein, B. J. Miller. Disc....798

Franksen, Ole Immanuel. Teaching the Uses of Digital Computers at the Technical University of Denmark. (60–1184)....781–7

Fraser, J. C.; G. I. Duncan, B. Valachovic. The Effects of Nuclear Radiation on the Electrical Strength of Air. (59–134)....19–26

Friedlaender, F. J. Disc....510

G

Gardner, Edward L. Instrumentation for Experimental Reactor Kinetic Studies. (60–514)..384–8

Gardner, F. H. An All-Transistorized Trunk Carrier System. (60–44)....212–16

Garrigus, John C. Design and Performance Characteristics of a High-Speed Wide-Chart Recorder. (60–113)....163–70

Gauster, W. F. Some Basic Concepts for Magnet Coil Design. (60–1409)....822–8

Getgen, L. E.; G. McGibbon. A New Control Terminal for Mobile Telephone Service. (60–159)...583–8

Geyger, William A. Miniaturized Current Transformers for Multirange Operation of Milliammeter Ink Recorders. (60–806)....681–9

Geyger, William A. Recent Developments on Magnetic-Coupled Multivibrators. (60–96)....106–12

Gilbert, Elmer G.; J. Otterman. The Synthesis of Linear Filters with Real or Imaginary Transfer Functions. (60–202)....323–30

Goudet, G. A 240-Line Fully Electronic Telephone Switchboard. (60–233)....232–41

Gould, L. A.; W. Kipiniak. Dynamic Optimization and Control of a Stirred-Tank Chemical Reactor. (60–982)....734–43; disc. 743

Graupe, K. K. The Analog Solution of Some Functional Analysis Problems. (60–1230)........793–7; disc. 798

Gray, P. E. The Effect of Source and Sink Thermal Resistance on Thermoelectric Generator Performance. (60–18)....15–19

Greenstein, J. A 2-Channel Data Link for Combined Analog-Digital Simulation. (60–10).........40–4

Griffith, C. G. A Transistorized Cable Carrier Repeater. (60–1251)....675–80

Guenther, J. H.; J. Shirman. Voice Immunity of Tone Operated Switching Centers. (60–820)........553–7

Gutzwiller, F. W.; T. P. Sylvan. Power Semiconductor Ratings Under Transient and Intermittent Loads. (60–69)....699–705; disc. 705

H

Hacker, M. E.; C. M. Davis. Automation in Switching Equipment Engineering. (60–778)........545–52

Hadley, Arthur H.; A. B. Rosenstein, G. Estrin. Disc.. 322

Hanna, D. A.; G. W. Marks. Quantitative Study of Low-Frequency Hysteresis Loops of Polarized Polycrystalline Barium Titanate. (60–1285).........799–808

Hanscom, F. E.; L. A. Jonas. High-Voltage Static Eliminators for Teletypewriters. (60–789).......576–8

Hansen, J. G. A Step-Driven Digital Magnetic Recorder. (60–183)....564–5

Harmuth, H. F. On the Transmission of Information by Orthogonal Time Functions. (79–1076).........248–55

Harmuth, H. F. Radio Communication with Orthogonal Time Functions. (60–142)..........221–8

Harple, K. G.; R. G. Lex, Jr. The LN 3000 Computer Control System. (60–215)....408–13

Hayter, S. G.; J. T. Wintermute. A Transistor-Oscillator Limit Switch for Indicating and Recording Instruments. (59–160)...........518–19

Hemmenway, S. F.; J. A. Russell, J. L. Scharf, P. C. Sharr. Magnetic Automatic Power-Range Control for an Aircraft Nuclear Reactor. (60–515)..........379–84

Hermach, F. L.; E. S. Williams. Thermal Voltage Converters for Accurate Voltage Measurements to 30 Megacycles Per Second. (60–135)..........200–05; disc. 206

Hernqvist, K. G.; F. H. Norman. The Empretron, a Mercury Pool Arc Tube Allowing Operation at Repetition Rates in the Kilocycle Range. (60–77) 85–91

Hesselgrave, D. H.; M. H. Kebby. The 57A Microwave Path Protection System. (60–1207)...654–9

Hoadley, Lloyd D.; H. E. Edgerton. Protection for Underwater Instruments. (60–662).......689–94

Hochgraf, Lester; E. J. Doyle, T. J. Talley, L. Pedersen, J. O. Smethurst. The E6 Telephone Repeater Background, Description, and Theory of Operation. (60–814)....525–42

Holcomb, P., Jr.; I. S. Coggeshall. A Character-Metered Transatlantic Switching System. (58–1143)....56–64

Hollander, G. L. Disc....793

Hoshall, C. H. A Power-Line Transient Recorder. (60–112)....170–3

Howell, A. S.; J. L. Faherty. A Military 4-Wire Field Telephone Incorporating an Electronic Signalling System. (59–259)....177–83

Howell, I. N.; J. P. Argo. Redstone Arsenal Closed-Circuit Educational Television System. (60–59).....185–90

J

Jalbert, B. W. An Analysis of the Operation of the Magnetic Second-Harmonic Modulator. (60–143) 268–72

Jamison, B. C.; H. E. Curtis, T. R. D. Collins. Interstitial Channels for Doubling TD-2 Radio System Capacity. (60–1248)....646–54

Jett, C. O. Wire Line and Carrier Communications on the Union Pacific Railroad. (60–815).....520–5

Johannessen, Paul R.; E. M. Lassiter, R. H. Spencer. High-Power Pulse Generation Using Semiconductors and Magnetic Cores. (60–870).......511–17

Jonas, L. A.; F. E. Hanscom. High-Voltage Static Eliminators for Teletypewriters. (60–789)..576–8
Jury, E. I. Disc..............................749

K

Kalbach, J. F. Disc..............................793
Kalman, R. E. Disc..............................743
Karkar, E. M.; R. W. Dean. Two-Wire Operation of N Carrier Systems. (60–998)..............594–7
Kebby, M. H.; D. H. Hesselgrave. The 57A Microwave Path Protection System. (60–1207)...654–9
Kent, Gordon. Generalized Brillouin Flow. (60–31)..............................144–8
Kipiniak, W.; L. A. Gould. Dynamic Optimization and Control of a Stirred-Tank Chemical Reactor. (60–982)..........................734–43; disc. 743
Kirkwood, L. W. Disc..............................322
Koenig, Eldo C. The Existence of an Electric Analog of a Magnetic Circuit Established by the Basic Laws Defining Magnetostatics. (56–118)..............................400–03
Konrad, G. T.; J. E. Rowe, H. W. Krage. Experiments on a Series of S-Band Crestatrons. (60–1411)..............................828–33
Krage, H. W.; J. E. Rowe, G. T. Konrad. Experiments on a Series of S-Band Crestatrons. (60–1411)..............................828–33
Krueger, Albert P.; J. C. Beckett. Experiments with Air Ions on the Trachea. (60–101)........695–9

L

Lassiter, Edward M.; P. R. Johannessen, R. H. Spencer. High-Power Pulse Generation Using Semiconductors and Magnetic Cores. (60–870)....511–17
Lee, Reuben; L. F. Deise, W. Etchison. Electronic Transformer Design by Digital Computers. (59–120)..........................314–22; disc. 322
Lehr, Samuel N.; V. P. Mathis. A Digital Start-up Control for Air-Cooled Nuclear Reactors. (60–512)..............................369–75
Lent, Worthington C. A Voice-Band Transmission-Measuring Set Using Swept-Band Techniques. (60–816)..............................570–6
LePage, W. R.; R. McFee. On the Response of Linear Systems to Periodic Excitations. (60–1288)..........................746–8; disc. 749
Lewis, R. F. Radio-Frequency Transmission Line Engineering. (60–899)....................634–41
Lex, R. G., Jr.; K. G. Harple. The LN 3000 Computer Control System. (60–215)................408–13
List, N. E.; L. W. Blumer, G. E. Reifenstuhl. Automation of Telephone Switching Equipment Engineering. (60–817)..............................560–4
Little, Edward M. Disc..............................593
Lord, Harold W. Disc..............................501
Love, B. E. Intermodulation Effects in FM and PM Systems. (60–161)..............................245–8
Ludwig, J. T. A Method for the Design of Holding Electromagnets..............................300–10
Ludwig, J. T. Inductors Biased with Permanent Magnets:
  Part I. (60–198)..............................273–8
  Part II. (60–199)..............................278–91
Luebbert, William F. Data Processing as a Tool for Generalizing Communications Systems. (59–1093)..............................431–6
Luft, Werner; E. J. Diebold. Transient Thermal Impedance of Semiconductor Devices. (60–68)..............................719–26
Luft, Werner; E. J. Diebold. Disc..............................705
Lush, Morley J. Disc..............................205
Lynn, G. E. Disc..............................482

M

MacDonald, D. L. Centralized Testing of a Very-High-Frequency Mobile Radio System. (60–996)..............................641–5
Maitland, Robert; B. B. Barrow. Transients in Logarithmic Count-Rate and Period Meters. (60–511)..............................362–9
Manteuffel, E. W. Time Response Constant and Transfer Function of Magnetic Amplifiers. (60–875)..............................473–82
Marks, G. W.; D. A. Hanna. Quantitative Study of Low-Frequency Hysteresis Loops of Polarized Polycrystalline Barium Titanate. (60–1285)..............................799–808
Mathis, Vernon P.; S. N. Lehr. A Digital Start-up Control for Air-Cooled Nuclear Reactors. (60–512)..............................369–75
Matick, Richard E.; G. W. Penney. Potentials in D-C Corona Fields. (60–33)................91–9
May, P. J.; T. M. Stump. Synthesis of a Resonant Transfer Filter as Applied to a Time Division Multiplex System. (60–822)..............615–20
McCluskey, E. J., Jr. Assignment of Carry Variables in Iterative Networks. (60–1211)..........772–8
McDuffie, G. E., Jr.; W. L. Chadwell. An Investigation of the Dynamic Switching Properties of 4-Layer Diodes. (60–19)..............................50–3
McFee, R.; W. R. LePage. On the Response of Linear Systems to Periodic Excitations. (60–1288)..........................746–8; disc. 749
McGibbon, Greg; L. E. Getgen. A New Control Terminal for Mobile Telephone Service. (60–159)..............................583–8
McMillan, M. B. Providing 150-Megacycle Telephone Service to North Carolina's Ocracoke Island. (60–793)..............................542–5
Meggitt, J. E. Error-Correcting Codes for Correcting Bursts of Errors. (60–654)................708–11
Meiksin, Z. H.; J. M. Walsh. Analysis and Optimization of the Bridge Magnetic Servoamplifier with True Motor Load. (60–784)..............413–20
Mertz, Pierre. The Effect of Delay Distortion on Data Transmission. (60–27)..................228–32
Meserve, W. E.; H. C. Torng. Disc..............339
Miller, B. J.; A. B. Clymer, T. F. Potts, W. C. Franke, C. R. Walli, G. N. Ornstein. Disc...........798
Mills, R. G. Fusion Power from Magnetically Confined Plasma. (60–1410)..................833–6
Mintzer, L.; M. Ringer. A Magnetic-Tape-to-Paper-Tape Converter. (59–1190)..............339–46
Moe, R. E. Tubes or Transistors: A Realistic Assessment. (60–73)..............................81–5
Moore, A. D. The Magnetic Analog with Spiral Search Coils. (60–129)................256–63; disc. 264
Myers, O. The New Nation-Wide Telephone Numbering Plan. (60–1210)..................673–5
Myles, A. H.; M. C. Davies, L. J. Srnka. D-C Magnetic Crane Hoist Control for A-C Powered Cranes. (60–30)..............................207–11

N

Nag, B. R. Locking Range of an Oscillator for Different Nonlinearities. (60–57)..................134–6
Neff, H. P.; J. D. Tillman. An Omnidirectional Circular Antenna Array Excited Parasitically by a Central Driven Element. (60–11)..........190–2
Nitzan, D.; H. C. Bourne. Rectifier Unblocking from a Generalized Analysis of Self-Saturating Magnetic Amplifiers. (60–91)..................99–105
Nixon, Floyd E. Disc..............................734
Norman, F. H.; K. G. Hernqvist. The Empretron, a Mercury Pool Arc Tube Allowing Operation at Repetition Rates in the Kilocycle Range. (60–77)..............................85–91
Nunn, R. Stewart. An Analysis of a Polarized Electromagnetic Drive System and of Vibrator Drive Systems. (60–201)..............................346–52

O

Okada, Robert H. Stable Transistor Wide-Band D-C Amplifiers. (60–6)..............................26–33
Ornstein, G. N.; A. B. Clymer, T. F. Potts, W. C. Franke, C. R. Walli, B. J. Miller. Disc......798
Ostendorf, B., Jr.; F. P. Pace. A Wired-Memory Translator with Shared Access. (60–126)..............................216–20
Otterman, Joseph; E. G. Gilbert. The Synthesis of Linear Filters with Real or Imaginary Transfer Functions. (60–202)..............................323–30
Otto, Robert E.; T. J. Williams. A Generalized Chemical Processing Model for the Investigation of Computer Control. (60–119)..........458–7

P

Pace, F. P.; B. Ostendorf, Jr. A Wired-Memory Translator with Shared Access. (60–126)..............................216–20
Parry, Charles A. Optimum Design Considerations for Radio Relays Utilizing the Tropospheric Scatter Mode of Propagation. (60–7)..............71–80
Pedersen, L.; E. J. Doyle, T. J. Talley, L. Hochgraf, J. O. Smethurst. The E6 Telephone Repeater Background, Description, and Theory of Operation. (60–814)..............................525–42
Penney, Gaylord W.; S. E. Craig. Sparkover as Influenced by Surface Conditions in D-C Corona. (60–103)..............................112–18
Penney, Gaylord W.; R. E. Matick. Potentials in D-C Corona Fields. (60–33)................91–9
Pennington, W. D.; R. W. Bratschi, A. L. Whitman. Teletypewriter Selective Calling Applied to Automatic Airlines Reservations. (60–861)..............................579–83
Pipes, Louis A. Difference Equations and Their Applications. (60–203)..........333–9; disc. 339
Poarch, C. K.; W. C. Ball. Telephone Influence Factor (TIF) and Its Measurement. (60–1195)..............................659–64
Potts, T. F.; A. B. Clymer, W. C. Franke, C. R. Walli, G. N. Ornstein, B. J. Miller. Disc............798
Preiss, Ralph J. Disc..............................793

R

Raillard, H. Fundamental Considerations of Power Dissipation Limits in Some Bistable Transistor Pulse Circuits. (60–20)..............................53–5
Rakshit, H.; N. N. Biswas. An Electronic Teleprinter Distributor. (60–770)..............................620–4
Reifenstuhl, G. E.; L. W. Blumer, N. E. List. Automation of Telephone Switching Equipment Engineering. (60–817)..............................560–4
Reinecke, Henry, Jr. Current-Operated Diode Logic Gates. (60–1188)..............................762–72
Reynolds, S. I. The Measurement and Influence of Surface Charges on the Inception Voltage Between Dielectric and Metal Dielectric Surfaces. (60–15)..............................310–14
Ringer, M.; L. Mintzer. A Magnetic-Tape-to-Paper-Tape Converter. (59–1190)..............339–46
Robinson, Paul B. A Precision, Continuous Voltage Reference for Industrial Recorders. (60–114)..............................197–200
Rogers, Walter E. Disc..............................263
Rosenstein, A. B.; G. Estrin, A. H. Hadley. Disc...322
Rowe, J. E.; G. T. Konrad, H. W. Krage. Experiments on a Series of S-Band Crestatrons. (60–1411)..............................828–33
Russell, J. A.; S. F. Hemmenway, J. L. Scharf, P. C. Sharr. Magnetic Automatic Power-Range Control for an Aircraft Nuclear Reactor. (60–515).379–84

S

Sandelien, J. F. An Approach to Dynamic Optimizing Control of the Continuous Process. (60–247)..............................291–9
Sattler, F. A.; L. C. Scala. Pigmented Polyester Magnet Wire for High-Temperature Operation. (59–1295)..............................503–05
Scala, L. C.; F. A. Sattler. Pigmented Polyester Magnet Wire for High-Temperature Operation. (59–1295)..............................503–05
Scharf, J. L.; J. A. Russell, S. F. Hemmenway, P. C. Sharr. Magnetic Automatic Power-Range Control for an Aircraft Nuclear Reactor. (60–515)..379–84
Scheneman, E. E.; S. K. Waldorf. Shall an Electron Tube or a Semiconductor Device Be Used? (60–74)..............................264–8
Schubert, E. J. Universal Modulus—Matrix Logic VI. (60–1181)..............................778–81
Schwab, Helmut. The Combi-System—A Proposal for New Concepts in Digital Data Processing. (60–158)..............................193–7
Schwartzenberg, J. W.; J. M. Finan, V. S. Underkoffler, R. N. Brey, Jr. A Multidecade Linear Reactor Simulator. (60–516)..............................353–8
Shackell, S. M.; J. G. Tryon. The Relative Merits of Incremental and Conventional Digital Computers in Air-Borne Real-Time Control. (60–120)..............................393–400
Sharr, P. C.; J. A. Russell, S. F. Hemmenway, J. L. Sharf. Magnetic Automatic Power-Range Control for an Aircraft Nuclear Reactor. (60–515).379–84
Shirman, J. A Transistor Automatic Gain Control System for Audio Frequencies. (60–813)...664–9
Shirman, J.; J. H. Guenther. Voice Immunity of Tone Operated Switching Centers. (60–820)..............................553–7
Sickles, Louis; J. A. Dever. A High-Speed Microcurrent Amplifier. (60–513)..............................375–9
Smethurst, J. O.; E. J. Doyle, T. J. Talley, L. Hochgraf, L. Pedersen. The E6 Telephone Repeater Background, Description, and Theory of Operation. (60–814)..............................525–42
Snyder, C. J.; C. A. Booker, Jr. A New Impulse Metering System. (60–64)..............................159–63
Spencer, Richard H.; E. M. Lassiter, P. R. Johannessen. High-Power Pulse Generation Using Semiconductors and Magnetic Cores. (60–870)....511–17
Srnka, L. J.; A. H. Myles, M. C. Davies. D-C Magnetic Crane Hoist Control for A-C Powered Cranes. (60–30)..............................207–11
Stapleton, C. A. Disc..............................624
Steiner, H. C.; R. W. Strecker. Voltage Transients Due to Arc Extinction. (60–76)..........139–44
Stoffels, R. E. Datacom Unit—An Experimental Data Transmission Subset. (60–790)..........669–73
Storm, H. F. Instabilities of Push-Pull Magnetic Amplifiers Feeding the Field of an Electric Machine. (56–730)..............................33–6
Storm, Herbert F.; C. W. Flairty. Theory of 3-Phase Bridge Magnetic Amplifier:
  Part II. (60–872)..............................451–8
  (Part I—1959)
Straiton, A. W.; C. W. Tolbert. Moon Reflection Studies with Bistatic Radar at 3,000 Mc. (60–518)..........................436–9; disc. 439
Strecker, R. W.; H. C. Steiner. Voltage Transients Due to Arc Extinction. (60–76)..........139–44
Stump, T. M.; P. J. May. Synthesis of a Resonant Transfer Filter as Applied to a Time Division Multiplex System. (60–822)..............615–20

Suzuki, T.; J. B. Thomas, T. R. Williams. Power Relationships and Temperature Dependence in the D-C Corona Field. (59–234)............1–4
Sylvan, T. P.; F. W. Gutzwiller. Power Semiconductor Ratings Under Transient and Intermittent Loads. (60–69)........................699–705; disc. 705
Sze, T. W.; C. W. Eggers. A New Function Generator. (60–1203)..............................756–62

T

Talley, T. J.; E. J. Doyle, L. Hochgraf, L. Pedersen, J. O. Smethurst. The E6 Telephone Repeater Background, Description, and Theory of Operation. (60–814)..............................525–42
Taylor, E. R.; A. E. Bachelet, C. A. Collins. New Television Network Switching Facilities. (60–811) 625–34
Thomas, J. B.; J. W. Drenning. Saturable-Reactor Control of Full-Wave and Biphase Rectifiers. (59–223)..................................36–9
Thomas, J. B.; T. R. Williams. Current-Voltage Relationships in Negative Pulsed Corona. (60–102)....................................136–9
Thomas, J. B.; T. R. Williams, T. Suzuki. Power Relationships and Temperature Dependence in the D-C Corona Field. (59–234)............1–4
Tillman, J. D.; H. P. Neff. An Omnidirectional Circular Antenna Array Excited Parasitically by a Central Driven Element. (60–11)........190–2
Tobin, W. H.; A. C. English. A Nondestructive Breakdown Phenomenon in Selenium Rectifiers. (58–821)......................................9–14
Tolbert, C. W.; A. W. Straiton. Moon Reflection Studies with Bistatic Radar at 3,000 Mc. (60–518) ....436–9; disc. 439
Torne, H. C.; W. E. Meserve. Disc...............339
Tryon, J. G.; S. M. Shackell. The Relative Merits of Incremental and Conventions; Digital Computers in Air-Borne Real-Time Control. (60–120)...... 393–400

U

Underkoffler, V. S.; J. W. Schwartzenberg, J. M. Finan, R. N. Brey, Jr. A Multidecade Linear Reactor Simulator. (60–516)..............353–8

V

Valachovic, B.; G. I. Duncan, J. C. Fraser. The Effects of Nuclear Radiation on the Electrical Strength of Air. (59–134)................19–26

W

Waldorf, S. K.; E. E. Scheneman. Shall an Electron Tube or a Semiconductor Device Be Used? (60–74)....................................264–8
Walli, C. R.; A. B. Clymer, T. F. Potts, W. C. Franke, G. N. Ornstein, B. J. Miller. Disc...........798
Walsh, J. M.; Z. H. Meiksin. Analysis and Optimization of the Bridge Magnetic Servoamplifier with True Motor Load. (60–784).............413–20
Warburton, C. R.; G. W. Altman, L. A. DeCampo. Automation of Computer Panel Wiring. (60–128) 118–25
Weed, Herman R. Portable Dynamic Electronic Speed-Torque Indicator. (60–660)........388–92
Whitman, A. L.; R. W. Bratschi, W. D. Pennington. Teletypewriter Selective Calling Applied to Automatic Airlines Reservations. (60–861)....579–83
Williams, E. S.; F. L. Hermach. Thermal Voltage Converters for Accurate Voltage Measurements to 30 Megacycles Per Second. (60–135)......... ....200–05; disc. 206
Williams, Theodore J.; R. E. Otto. A Generalized Chemical Processing Model for the Investigation of Computer Control. (60–119)..........458–73
Williams, T. R.; J. B. Thomas. Current-Voltage Relationships in Negative Pulsed Corona. (60–102)....................................136–9
Williams, T. R.; J. B. Thomas, T. Suzuki. Power Relationships and Temperature Dependence in the D-C Corona Field. (59–234)...........1–4
Wintermute, J. T.; S. G. Hayter. A Transistor-Oscillator Limit Switch for Indicating and Recording Instruments. (59–160)...........518–19
Wolf, Alfred A. Analysis of Transcendental Nonlinear Systems. (60–879)......................449–51
Wolff, G. Back-Transient Diode Logic. (59–261)... 4–9

Y

Yamakura, S. Disc...........................837
Yeh, Leang P. Use of Moon or Satellite Relays for Global Communications. (60–1064).......607–15

Z

Zemanian, Armen H. The Properties of Pole and Zero Locations for Nondecreasing Step Responses. (60–782)................................421–6

Vol.
Pt.
196

Power Relationships and Temperature Dependence in the D-C Corona Field

J. B. THOMAS
MEMBER AIEE

T. R. WILLIAMS
NONMEMBER AIEE

T. SUZUKI
NONMEMBER AIEE

THE STUDY of d-c corona discharge over a period of many years has resulted in a considerable body of literature on the subject. However, relatively little has been reported on corona properties over extended temperature ranges. At least two experimental studies have been published, both using coaxial cylindrical geometry. The first, by Koller and Fremont,[1] treats negative corona at temperatures to 1,400 R (degrees Rankin) and pressures to 5 atmospheres. The second, by Thomas and Wong,[2] treats both positive and negative corona at temperatures to 2,000 R and pressures to 8 atmospheres.

This paper is, in some ways, an extension of reference 2 and was undertaken in an attempt to determine two effects: first, the effect of heating only the discharge electrode in a coaxial cylindrical arrangement and, second, the effect of the voltage waveform on the d-c power dissipated in the corona field.

Apparatus and Procedure

The test cell and associated equipment are shown schematically in Fig. 1. The cell itself consisted of a vertical aluminum tube 12 inches long and 0.5 inch in inside diameter, and a coaxial nichrome wire 14 inches long and 0.010 inch in diameter. The ends of the tube were flared to prevent field concentrations. The wire was attached to a heavier bus wire as shown and was held taut by brass weights. Porcelain insulators supported the tube, the wire, and a weight guide structure which facilitated the centering of the wire in the tube.

The wire (or discharge electrode) was heated by having passed through it a current obtained from an isolating heater transformer supplied by a variable 60-cycle source. The temperature of the wire was determined by means of measuring its linear expansion with the aid of a simple ratio arm which served as a mechanical amplifier.

The average current flow to the tube (or collecting electrode) was measured by means of a d-c milliammeter connected between the tube and ground while effective voltage across the corona gap was determined by an electrostatic voltmeter.

Three types of voltage waveforms, as shown in Fig. 2, were used to energize the test cell. That of Fig. 2(A) was a flat d-c voltage with a ripple less than 1.0%. That of Fig. 2(B) was obtained by the full-wave rectification of a 60-cycle per second voltage. In this case the power supply was heavily loaded with a bleeder resistor to prevent capacitive filtering by the test cell and to retain essentially the shape shown. In the third case a pulse generator was used to provide the energizing voltage. The pulses were each one microsecond in width and were repeated at a rate of 5,000 per second. The basic pulse shape from the generator is shown in Fig. 3. A blocking diode was added in series with the generator and test cell to prevent the test cell from discharging below corona onset voltage. The resulting voltage applied to the test cell is shown in Fig. 2(C).

Both positive and negative polarities of the three waveforms just discussed were used. The blocking diode for the pulse generator is shown in Fig. 1 for the negative pulse case and was, of course, reversed for positive pulses.

Also given in Fig. 2 are the approximate corresponding current waveforms. They are all drawn for high current densities; the small random components are not shown. Fig. 3 shows in more detail a typical current pulse for the pulse-energized case.

The basic data taken were average current through the test cell, effective voltage across the cell, and wire temperature. In all cases the test cell was operated in air at atmospheric pressure and room temperature.

Considerable care had to be exercised to obtain reproducible data. The wire used had a rather small diameter and was easily damaged. It was cleaned with carbon tetrachloride before mounting, then operated for approximately one hour at high negative corona currents which tended to "clean up" the surface and stabilize discharge. The wire was replaced frequently since it tended to be permanently stretched, particularly after operating at high temperatures. This stretching reduced wire diameter and increased discharge currents.

When current-voltage readings were taken for a fixed heater current to the wire, it was found that the wire temperature varied greatly over the range of discharge current values. Specifically, the wire temperature decreased rapidly as soon as corona current began. As the corona current was increased, the wire temperature reached a minimum and then increased slightly. The total temperature change was as much as 300 F (degrees Fahrenheit) at the higher temperatures. It was presumed that the initial large temperature decrease was due to air convection currents caused by the corona discharge and that the final slight increase in temperature resulted from the heating of the air space by corona losses. Although the results were the same in both cases, it was found more convenient, particularly at higher temperatures, to hold the wire temperature constant by varying the heater current rather than operating at a fixed heater current.

Current-Voltage Characteristic

Current-voltage characteristics were obtained for the three voltage waveforms of Fig. 2 for both positive and negative polarity; thus six cases were studied. The data from these are presented in Figs. 4(A), 5(A), 6(A), 7(A), 8(A), and 9. It was convenient to plot voltage versus relative air density, with corona current as

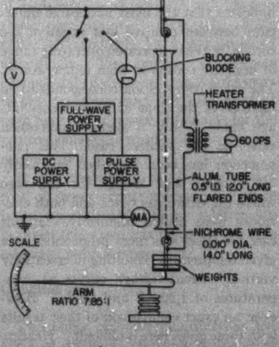


Fig. 1. Diagram of apparatus

Paper 59-234, recommended by the AIEE Electronics Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 1-6, 1959. Manuscript submitted November 6, 1958; made available for printing September 17, 1959.

J. B. Thomas, T. R. Williams, and T. Suzuki are with Princeton University, Princeton, N. J.

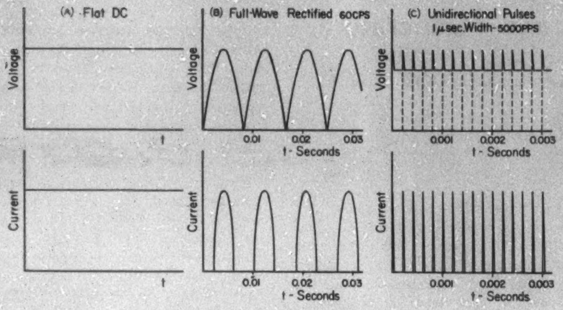


Fig. 2. Voltage and current waveforms

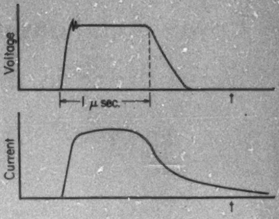


Fig. 3. Detail of pulse shapes

a parameter. The density scale used was chosen for convenience and unit density was taken to be that at 1,000 R. The relative density range shown corresponds to a temperature range of approximately 70 to 1,200 F.

The current-voltage characteristics were obtained over as wide a voltage range as possible, the lower limit of the range being determined by corona onset, and the upper by sparkover. The voltage sparkover values are not given; they correspond roughly to the upper curve in each of the six figures previously mentioned, i.e., curves are given for as high a current as could be obtained consistently without sparkover. The 10-microampere curves can be used as a measure of corona onset voltage, and thus the vertical distance between the lower and upper curve for each of the figures corresponds to the voltage operating range of the cell as a corona device for a given temperature and type of energization.

Flat D-C Energization

A comparison of Figs. 4(A) and 5(A) indicates that the most noticeable difference between positive and negative d-c corona is the compressed operating range of negative corona at high temperatures. The range compression corresponds to an increase in the slope of the current-voltage curve. Only a slight compression exists in the positive case. This phenomenon has been described previously by Thomas and Wong[2] for the case where the entire gas space and both electrodes were heated uniformly. In this case the effect was more pronounced at a given temperature, resulting in essentially vertical current-voltage curves at temperatures of 1,200 F and above. However, an exact comparison of their results with those given here is difficult. Heating the wire alone insures that the gas immediately surrounding the wire is also heated to approximately the same temperature. However, the temperature gradient across the corona gap is quite large since the outer electrode is at a temperature roughly equal to room temperature. The temperature gradient is also unknown since it appears to be affected by corona current.

Two other points should be mentioned briefly in the comparison of positive and negative d-c corona. The curves of Figs. 4(A) and 5(A) are not monotonic but contain points of inflection. It seems reasonable to assume that these are the result of the convection effects previously discussed. In addition, the well-known phenomenon[3,4] should be noted that the corona-starting voltages are not as well-behaved in the negative case as in the positive.

Full-Wave Energization

As might be expected, a lower effective voltage is required for the full-wave case

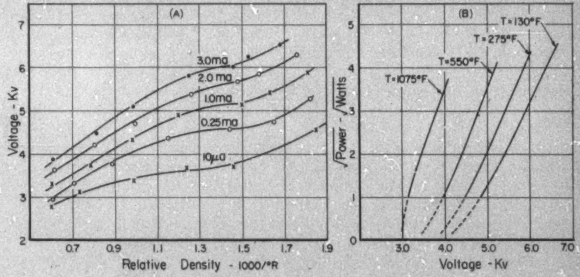


Fig. 4. Negative corona characteristics, flat d-c voltage

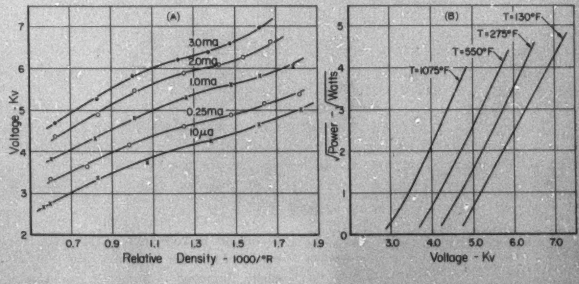


Fig. 5. Positive corona characteristics, flat d-c voltage

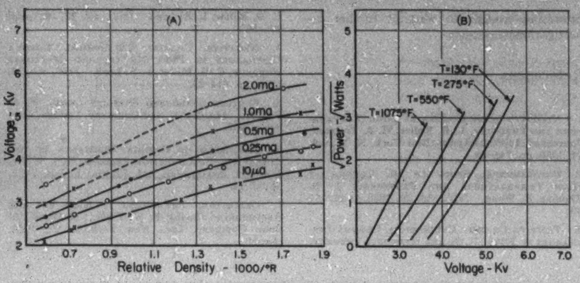


Fig. 6. Negative corona characteristics; full-wave d-c voltage

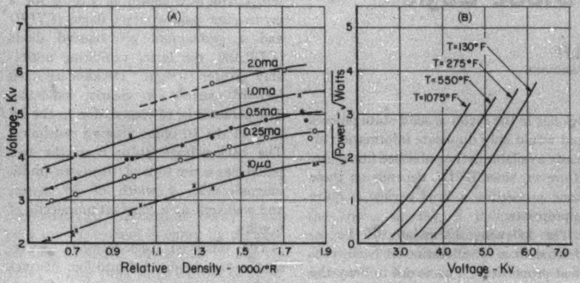


Fig. 7. Positive corona characteristics; full-wave d-c voltage

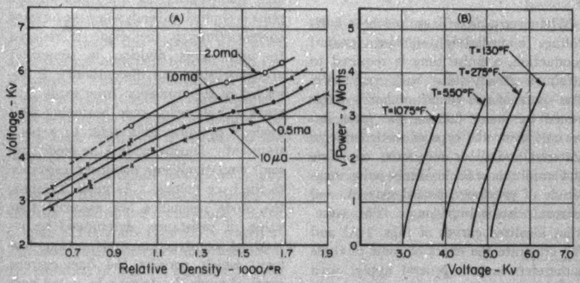


Fig. 8. Negative corona characteristics; pulsed voltage

for the same current. Hence the curves of Figs. 6(A) and 7(A) are displaced downward compared with those of Figs. 4(A) and 5(A). This is a result of the nonlinear corona current-voltage relationship. The same voltage range compression may be seen in Fig. 6(A) for the negative full-wave case as in Fig. 4(A) for negative flat direct current.

In addition, sparkover occurs at a lower effective voltage than before, since it is determined by the peak value of the voltage. As a consequence, the maximum currents obtainable are lower than for the previous case.

Parts of some of the curves of Figs. 6(A) and 7(A) are shown as broken. Over these regions the discharge wire tended to vibrate, and sporadic arcing occurred. Sufficient weight could not be added to stop the vibration without exceeding the elastic limit of the wire.

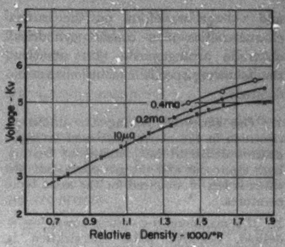


Fig. 9. Positive corona characteristics; pulsed voltage

PULSE ENERGIZATION

The positive and negative pulse corona characteristics differ radically from each other. Negative pulse data are given in Fig. 8(A); they are quite similar to the negative flat direct current data except that the voltage range is considerably compressed at all temperatures and maximum currents obtainable are lower. The reasonably high voltages and currents are all the more interesting when the short pulse length and low repetition rate are considered. No attempt was made to determine optimum pulse lengths and repetition rates.

Positive pulse characteristics were limited by sparkover to a very narrow range, as shown in Fig. 9. A comparison with Fig. 5(A) shows that the curves which were obtained are similar to those of positive flat direct current.

Power Relationships

One of the purposes of this study was to obtain a measure of the d-c power dissipated in the various corona fields just discussed. This dissipation is important in such applications as the design of high-voltage regulators[5] and electrostatic precipitators.[6] Figs. 4(B), 5(B), 6(B), 7(B), and 8(B) have been prepared accordingly. The product of average current and effective voltage is a measure of the d-c power and has been used as such here. The power obtainable for the positive pulse case was too low to warrant consideration.

It was convenient to plot the square root of power versus voltage since there is an approximately linear relationship between these two quantities for positive corona at room temperature.[7]

Conclusions

This paper presents quantitatively the effect of discharge electrode temperature and voltage waveform on the corona characteristics of a coaxial cylindrical electrode arrangement. In addition, the following specific conclusions are indicated:

1. The effect of heating the discharge electrode in the above-mentioned arrangement is similar to that obtained by heating both electrodes and the gas space, but the effect is less pronounced for the same temperatures.

2. If a maximum d-c corona power loss is desired, negative pulse energization is promising enough to warrant further investigation.

References

1. NEGATIVE WIRE CORONA AT HIGH TEMPERATURE AND PRESSURE, L. R. Koller, H. A. Fremont. *Journal of Applied Physics*, New York, N. Y., vol. 21, 1950, pp. 741–44.

2. EXPERIMENTAL STUDY OF DC CORONA AT HIGH TEMPERATURES AND PRESSURES, J. B. Thomas, E. Wong. *Ibid.*, vol. 29, 1958, pp. 1226–30.

3. POSITIVE COAXIAL CYLINDRICAL CORONA DISCHARGES IN PURE N_2, O_2, AND MIXTURES THEREOF, C. G. Miller, L. B. Loeb. *Ibid.*, vol. 22, 1951, pp. 494–503.

4. NEGATIVE COAXIAL CYLINDRICAL CORONA DISCHARGES IN PURE N_2, O_2, AND MIXTURES THEREOF, C. G. Miller, L. B. Loeb. *Ibid.*, vol. 22, 1951, pp. 614–21.

5. VOLTAGE STABILIZED SUPPLIES (book), F. A. Benson. Macdonald and Company, London, England, 1957.

6. THE ROLE OF CORONA DISCHARGE IN THE ELECTRICAL PRECIPITATION PROCESS, H. J. White. *Electrical Engineering*, vol. 71, no. 1, Jan. 1952, pp. 67–73.

7. DIELECTRIC PHENOMENA IN HIGH VOLTAGE ENGINEERING (book), F. W. Peek. McGraw-Hill Book Company, Inc., New York, N. Y., 1929, chap. 6.

Back-Transient Diode Logic

G. WOLFF
ASSOCIATE MEMBER AIEE

CONVENTIONAL diode switching circuits show a significant decline in performance and efficiency as their speed of operation is increased. Faster rise and fall times require correspondingly greater peak currents to charge and discharge circuit capacitance, at the cost of increased static power dissipation, drive requirements, and noise. These peak currents are in demand for only a brief initial and final fraction of the total signal-time duration and might ideally be gated by a 2-step switch having a low resistance during these brief time intervals, with a much greater resistance for the remainder of the signal duration. The semiconductor diode approximates this function when rapidly switched in a low impedance circuit immediately following forward conduction. This is due to the well-known hole-storage effect, a property of diodes generally considered detrimental for rapid switching operations. However, it will be shown that hole storage may also serve as the design basis for a new type of diode-switching circuit, henceforth referred to as back-transient diode logic (BTDL), which exhibits good efficiency, low noise, and excellent compatibility with vacuum tubes or transistors at switching speeds as high as 30 mp (megapulses) per second.

Diode Transient Properties

The forward and reverse transients associated with semiconductor diodes have been the subject of extensive analysis in the past 10 years, ranging from highly theoretical treatises to practical considerations of devising suitable test methods.[1–8] A full discussion properly belongs in the realm of solid-state physics and would only duplicate information already available in the published literature. However, since BTDL depends on these same properties, a brief review is quite appropriate.

The following discussion will be restricted to a consideration of back-transient properties. This is not to deny the importance of forward-transient effects, but their presence does not significantly affect the operation of the circuit to be discussed.

Whenever a diode is pulsed by a back voltage immediately following forward conduction, a finite time is required to attain a steady-state back resistance. The shape of this time function is determined by many factors, the most important being the type of semiconductor material and diode structure, rise time and amplitude of back-voltage pulse, magnitude of previous forward current, and external circuit impedance. The somewhat idealized curves of Figs. 1(A) and (B) illustrate the effect of some of these parameters and in general apply, with only a change of scales, to all types of semiconductor diodes presently available.

The simple test circuit of Fig. 2 was constructed to observe the back-transient current of a given diode with the aid of a high-speed oscilloscope. The current viewing resistance as well as the internal impedance of the pulser were kept as small as was practically possible. It will be shown later that this corresponds to circuit conditions encountered in high-speed diode logic circuits. Voltage *V* can be adjusted for any desired forward current through the test diode. Some typical waveshapes obtained with this test circuit are shown in Fig. 3 for a germanium point-contact diode (*IN191*) and a germanium gold-bonded diode (*CTP309*), the latter exhibiting noticeably greater storage. Dependence upon forward current is clearly indicated. The pedestal is, of course, due to the *IR* drop in the 10 ohms viewing resistance. The back-voltage pulse from the test pulser has a rise time of 10 mμsec (millimicroseconds), a width of 70 mμsec, and was kept at a constant amplitude of 6 volts.

The observed back-transient waveshapes may be divided into four distinct modes of operation, indicated in Fig. 4. Mode *1* represents the steady-state forward conduction period. Mode *2* begins immediately as the back-voltage pulse is applied and ends when the forward current through the diode is reduced to zero. This division is somewhat arbitrary since no discontinuity is involved in the transition from mode *2* to mode *3*. However, only beginning with mode *3* can one truly refer to a back transient. During mode *2* the diode presents a relatively low, nonlinear resistance to the back voltage. As mode *3* begins, the diode resistance has assumed that value of resistance, determined by its material, construction, and previous history, which permits the back-current surge of mode *3*. It may be noted here that two distinct time constants are indicated by modes *3* and *4*, whereas a single

---

Paper 59-261, recommended by the AIEE Electronics Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 1–6, 1959. Manuscript submitted November 3, 1958; made available for printing September 23, 1959.

G. WOLFF is with Denver Research Institute of the University of Denver, Denver, Colorado.

This paper is a digest of a thesis submitted in partial fulfillment of the requirements for a Masters Degree in Electrical Engineering at Polytechnic Institute of Brooklyn, Brooklyn, N. Y. The work was performed as part of a research project sponsored by the U. S. Department of Defense.

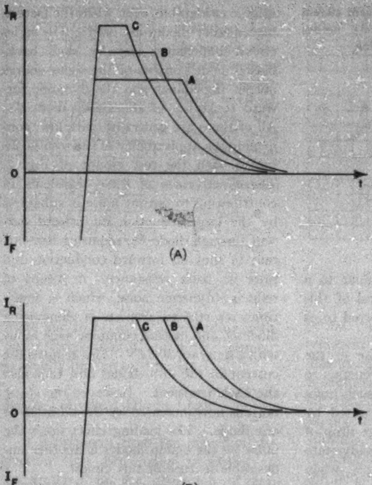


Fig. 1. Back-transient current characteristics

A—As function of series resistance or back voltage. $R_A > R_B > R_C$, V constant; $V_C > V_B > V_A$, R constant

B—As function of forward current $I_A > I_B > I_C$, R and V constant

l. 79
1
60

time-constant exponential decay from peak reverse current is shown by the previously discussed waveforms of Fig. 1. However, the peak-surge current of mode *3* is clearly evident in the recorded observations of Fig. 3 and agrees with the findings of at least one investigator.[8] Larger values of series resistance and slower oscilloscopes tend to mask this effect; perhaps this explains why little attention has been paid to it so far in the literature. Thus mode *3* actually represents the major part of the back transient since even at the beginning of mode *4*, the diode back resistance is already in the order of thousands of ohms.

A severe limitation in experimentally observing these transients is imposed by the rise time of the oscilloscope. Even though a Tektronix type *517A* (rated at 7 mμsec maximum rise time) was used, the duration of mode *3* was shorter than the traces of Fig. 3 indicate. More accurate results are obtainable by means of very wide bandwidth oscilloscopes, such as the Edgerton, Germeshausen, & Grier travelling-wave tube type, but this is not a necessity for relative measurements of a practical nature.

High-Speed Considerations of Diode Switching Circuits

The general operation and design principles of diode switching circuits have been extensively described,[9–13] and the following brief discussion will be restricted to problems of high-speed operation, including the effects of the diode transient properties described previously.

A typical AND gate having *n* inputs is shown in Fig. 5. All voltages are referred to ground, *C* includes the capacitance of stray wiring, diode D_s, and connected load. Three considerations are of main concern:

1. The rise time of the output signal when all inputs are present.
2. The fall time of the output signal when one or more inputs are removed.
3. The output noise when all inputs but one are present.

In order for the gate to function correctly, the following conditions must be satisfied:

$$V_1 > V_3 > V_2, \quad I_2 > I_1$$

When all inputs are present and assuming ideal diodes, the output voltage, e_0, will rise exponentially as follows:

$$e_0 = (V_1 - V_3)(1 - e^{-t/R_1C}) + V_3 \qquad (1)$$

In practical designs the voltage drop across R_1 is made to be much greater than the final output amplitude so that I_1 may be assumed constant. In that case the output voltage can be expressed by the simple relationship:

$$e_0 = I_1 t_r / C \qquad (2)$$

where t_r = rise time of e_0.

The fall time of the output voltage is governed by similar considerations with V_2 and R_2 as added parameters. Also e_0 cannot exceed in amplitude any of the inputs present.

In order to increase switching, speed must be decreased and I_1 correspondingly increased to maintain a constant output. For rise times of several mμsecs, currents in the order of tens of milliamperes (ma) and resistors dissipating watts are not unusual. Unfortunately, this power is to a very large degree wasted since it is required for only a brief initial fraction of the total signal-time duration.

It has been noted previously that one

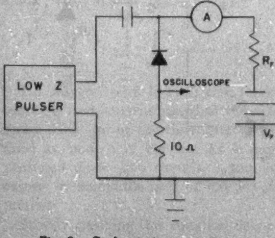


Fig. 2. Back-transient test circuit

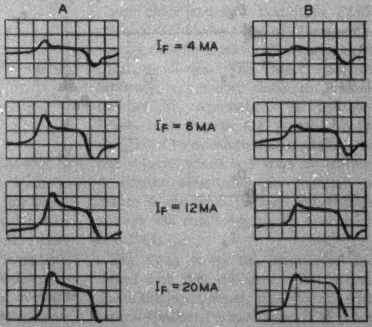


Fig. 3. Typical oscillograms of back-transient currents. Vertical, 10 ma per division; horizontal, 20 mμsec per division

A—Diode type CTP 309
B—Diode type IN191

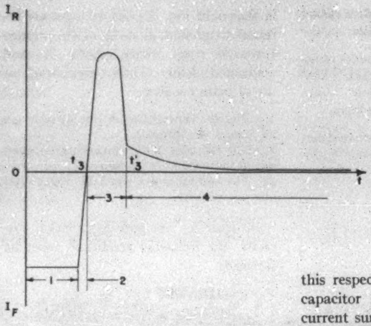


Fig. 4. Back-transient current characteristic showing modes of operation

of the conditions which must be satisfied is that $I_2 > I_1$. As I_1 increases with speed, I_2 does likewise, and this creates several problems. Thus the number of inputs that can be driven from a given source is successively reduced as switching speed goes up. Still more troublesome is the effect of this heavier current upon the conducting diodes. We have seen that the back transient is directly proportional to the forward current passed by the diode prior to being back-biased. The clamp diode, D_c, carries the heaviest load. Its current is defined by:

$$I_3 = nI_2 - I_1 \qquad (3)$$

Therefore a large back-current surge is passed by diode D_c before it can attain a reasonably high back resistance. In this respect its behavior is similar to a capacitor and the time integral of this current surge is sometimes referred to as "lost charge."

The problem of output noise in the presence of all but one input remains to be considered. The origin of such noise is twofold. Static noise is caused by variation in the forward voltage drop of diode D_c due to changes in steady-state current conditions. For example, when only one input remains unpulsed, I_3 decreases from its quiescent value of $nI_2 - I_1$ to $I_2 - I_1$. This is often a substantial change and may give rise to an output ranging from several tenths of a volt to almost a volt, depending upon the type of diode chosen for D_c. Dynamic noise is usually the more serious source of unwanted output at high switching speeds. It is the result of back transients passed by n-1 diodes, assuming worst condition, plus the effect of any parallel capacitance. The currents thus passed may be quite considerable, even when the fastest, high-quality diodes are used. The series circuit impedance seen by each back-biased diode consists of the signal-source output impedance and the dynamic forward resistance of the clamp diode D_c. All of these are generally quite low, confirming the applicability of the results obtained with the test circuit of Fig. 2. The effectiveness of these transients in contributing to output noise is enhanced by the large reduction in forward current through diode D_c, requiring less current to shut off forward conduction and raise its back resistance. A means of reducing dynamic noise, which is sometimes resorted to, consists of eliminating diode D_c and instead clamping each input with a separate diode.[14] This reduces the current in the AND diodes and thus also the back transient. However the static noise is now increased by the addition of one diode. The loading effect upon the pulse by the clamp diodes is another undesirable feature of this circuit.

OR gates present much less of a problem than AND gates at high switching speed. OR diodes should be chosen primarily for optimum forward transient properties. Their back transients are usually not deleterious; on the contrary, they often serve as a means of obtaining a more rapid discharge of circuit capacitance at signal termination.

As switching circuits are successively cascaded, impedance levels decrease rapidly.[12] Since the initial resistances of high-speed diode switching circuits are

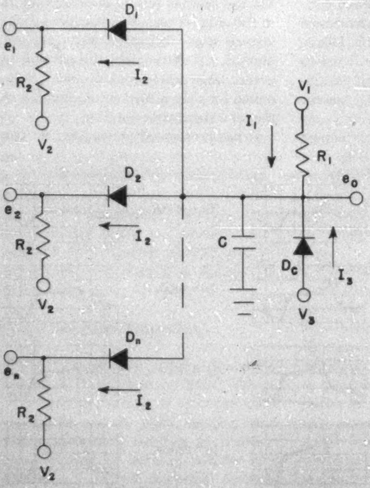


Fig. 5 (left). Conventional diode AND gate

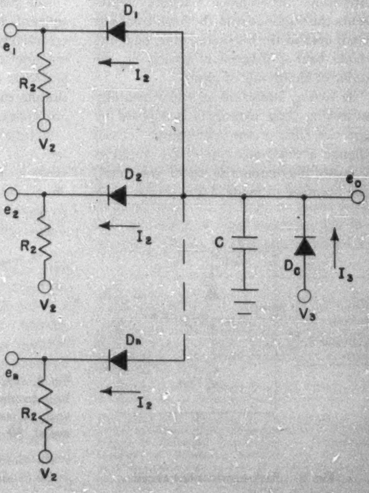


Fig. 6 (right). BTDL AND gate

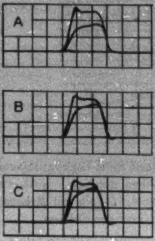


Fig. 7. BTDL AND gate waveshapes with number of inputs as variable parameter. C = 10 μμf, I_2 = 2 ma. Vertical, 2.5 volts per division; horizontal, 40 mμsec per division

A—2 inputs
B—3 inputs
C—4 inputs

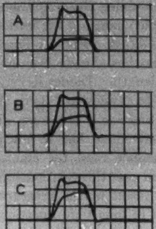


Fig. 8. BTDL AND gate waveshapes with I_2 as variable parameter. C=10 μμf, 2 inputs. Vertical, 2.5 volts per division; horizontal, 40 mμsecs per division

A—I_2=0.25 ma
B—I_2=1 ma
C—I_2=2.5 ma

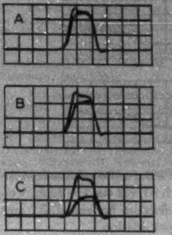


Fig. 9. BTDL AND gate waveshapes with C as variable parameter. I_2=3 ma, 4 inputs. Vertical, 2.5 volts per division; horizontal, 40 mμsecs per division

A—C=5 μμf
B—C=10 μμf
C—C=25 μμf

quite low to begin with, cascading without interposition of an active element is usually limited to three levels.

Back-Transient Diode Logic

It is apparent from the preceding discussion that two major problems are encountered in the operation of conventional diode switching circuits at high speeds: 1. greatly increased quiescent power dissipation, and 2. undesirable effects of diode transient properties.

A new type of diode switching circuit will now be described for which quiescent power dissipation is not dependent upon speed of operation. Since diode back transients are an essential element of these circuits, providing low-impedance paths for charge and discharge currents, the name back-transient diode logic is an apt designation.

Fig. 6 represents an AND circuit which upon closer examination is seen to differ from the one of Fig. 5 only by the absence of V_1 and R_1, nevertheless this is sufficient to result in radically different circuit operation. The following steady-state conditions must be satisfied:

$$I_3=nI_2,\quad V_3>V_2$$

When all inputs are pulsed, the output voltage, e_0, is now determined by the following equation:

$$e_0=\left(\int_{t_3}^{t_3'} i_{D1}\,dt+\int_{t_3}^{t_3'} i_{D2}dt+\dots \int_{t_3}^{t_3'} i_{Dn}dt-\int_{t_3}^{t_3'} i_{Dc}dt\right)1/C \qquad (4)$$

The integral term in each case represents the back current passed by the subscript diode during the time interval defined as mode *3* in Fig. 4. As in conventional diode AND circuits, e_0 cannot exceed in amplitude any of the inputs present.

Strictly speaking this equation defines e_0 for only the initial portion of the signal period. During this time the AND diodes represent a low resistance, in order of magnitude ranging perhaps from tens to hundreds of ohms, and the current passing through them must be sufficient to supply the back current of diode D_c in addition to charging the parallel circuit capacitance to the desired voltage. For the remainder of the signal duration, e_0 is determined by the diode characteristics of mode *4*. The magnitude of time constants during this mode causes e_0 to change very slowly compared to mode *3*. The effect upon signals of long-time duration will be considered later. In pulse-type synchronous system of logic, clock rates of 10 mp per second and higher correspond to pulse widths of 50 mμsec and less. For such short-time intervals mode *3* is essentially the sole determinant of output amplitude.

When all inputs but one are pulsed, the expression for the output noise voltage becomes:

$$e_0=\left(\int_{t_3}^{t_3'} i_{D1}dt+\int_{t_3}^{t_3'} i_{D2}dt+\dots \int_{t_3}^{t_3'} i_{Dn-1}dt-\int_{t_3}^{t_3'} (i_{Dc}+I_2)dt\right)1/C \qquad (5)$$

The total charge available to produce an output has now been reduced by

$$\int_{t_3}^{t_3'} i_{D_n}\,dt$$

At the same time the charge demand has been increased by

$$\int_{t_3}^{t_3'} I_2dt$$

This combination makes it possible, if parameters have been correctly chosen, to keep dynamic noise at very low levels and still obtain the desired output voltage when all inputs are present. Since the integral terms of these equations are, for the particular diode concerned, primarily functions of forward current, the choice of diodes and currents determines the design of BTDL circuits to a large extent.

The fall time of e_0 is not necessarily governed by the value of R_2. In general the inputs are coupled through OR diodes whose back transients can be utilized for low-impedance discharge paths. The output will then follow the signal source, as it also would do in the case of capacitor coupled inputs.

Data obtained by means of the test circuit of Fig. 2 have proved very useful in designing practical BTDL circuits. The area under the back-transient current trace during mode *3* is obtained most easily by assuming a triangular shape. With a properly calibrated oscilloscope this permits a very rapid determination of a diode's storage capability at any desired value of forward current. The results are sufficiently accurate for a preliminary design which may then be empirically optimized. As a simple example, consider a 2-input AND gate which is to have an output of 3 volts across the clamp diode in parallel with a capacitance of 10 μμf (micromicrofarads). This requires that 30×10^{-12} coulombs be supplied to charge the capacitance. A Hughes *HD2109* point-contact diode (30×10^{-12} coulombs storage for 4-ma forward current) was selected as the clamp diode and two Transitron *T15G* gold-bonded diodes (30×10^{-12} coulombs storage each for 2-ma forward current) as the AND diodes. The choice is quite arbitrary, of course, since an infinite number of combinations is possible. When both inputs are pulsed, apply equation 4 to calculate the resultant output voltage, which is 3 volts as desired. A similar substitution in equation 5 makes it apparent that no net output voltage will result when only one input is pulsed.

Fig. 7(A) shows the waveforms obtained with this AND gate. The largest one is the input signal, capacitor coupled from a Hewlett-Packard type *212A* pulse

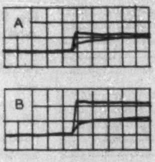


Fig. 10. **Effect of input pulse amplitude variation upon input. $I_2 = I_3 = 2$ ma, $C = 8 \mu\mu f$. Vertical, 10 volts per division, horizontal, 0.1 μsec per division**

A—11-volt input
B—22-volt input

generator. Its repetition rate is 5 kc. The output pulse is superimposed and its leading edge can be seen to rise to an amplitude of from 3 to 4 volts. During the flat-top portion of the input pulse, the output continues to increase, though at a much slower rate, being now determined by conditions of mode *4*. The ratio of parallel static back resistances of the AND diodes to the static back resistance of the clamp diode determines the final value of output voltage, assuming the input signal continues long enough for static conditions to be reached. The output with only one input signal present is shown by the virtually straight line at the base of the pulses, indicating negligible noise.

The addition of inputs using similar diodes, with the forward current in each maintained at 2 ma, will cause an increase in output voltage. Since there is a corresponding increase in current through the clamp diode, the net output voltage will not necessarily vary in direct proportion to the number of inputs. Figs. 7(B) and (C) show input, output, and noise voltage waveshapes for a 3- and 4-input AND gate respectively. The steeper rise of the output voltage during the flat-top portion of the input pulse as more inputs are added reflects the resultant lowering of the parallel AND diode resistance.

So far the number of inputs has been the only parameter which was varied. Figs. 8 and 9 clearly show the dependence of output voltage upon forward current and output capacitance. Again the larger pulse always represents the input signal, the superimposed smaller pulse the output voltage. The noise voltage is not included in these pictures. As in Fig. 7 it is insignificant.

Further insight into some aspects of BTDL may be gained by eliminating all but one of the inputs of the circuit shown in Fig. 6. For example, Figs. 10(A) and (B) show the influence of input-voltage amplitude variation upon the output. It can be seen that doubling this amplitude causes no noticeable change in the initial value of output voltage. Fig. 11 illustrates an effect that has already been referred to at some length. Diodes having similar transient properties are used in the same relative position. However, in one case the input diode has a much higher static resistance than the clamp diode, in the second case this relationship is reversed. The result in the presence of an input signal of long-time duration (10 μsec) is clearly evident. The larger pulse is the input and its noticeable droop is caused by capacitor coupling. The ratio of input to output voltage at the trailing edge of the pulse in each case closely corresponds to the ratio of measured static back-resistance values of the two diodes concerned. An indication is thus given that BTDL does not necessarily have to be restricted to very high speeds or clocked systems.

Several considerations of a more practical nature remain to be examined. In view of the temperature dependence of diode transient properties, circuits based upon these properties might well be expected to exhibit a similar dependence. However, BTDL is relatively temperature insensitive because variations in the AND diodes tend to be compensated by similar variations in the clamp diode.

The problem of diode selection and matching becomes very important when large quantities are required. Measurement of the peak amplitude of mode *3* at only one value of forward current is adequate for production purposes and may be performed very rapidly by means of the test circuit of Fig. 2.

Under actual operating conditions, noncoincident arrival of signals at a multi-input switching circuit is commonly encountered. In addition, one or more of these signals may be static rather than dynamic. BTDL AND gates will respond correctly and no output will appear until the last input is present. However, this does not imply that there will not be some difference in output amplitude when all inputs appear either simultaneously or successively. Allowance for this variation can be made in the original design, especially in view of the fact that signals are often limited by clamping to a predetermined level.

BTDL circuitry must, of course, be combined with active elements such as vacuum tubes or transistors in any but the most elementary logic elements. Vacuum tubes require rapid charge and discharge of their grid capacitance when used as high-speed switches and the suitability of BTDL is immediately apparent. The output from BTDL gates is likewise an excellent base drive for fast switching transistors, which demand high forward and reverse-current surges at the start and finish of pulse inputs.

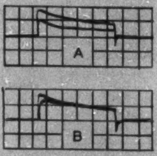


Fig. 11. **Effect of diode static-back resistances with pulses of long-time duration. $I_2 = I_3 = 2$ ma, $C = 8 \mu\mu f$. Vertical, 10 volts per division, horizontal, 2 μsec per division**

A—Low back-resistance clamp diode
B—High back-resistance clamp diode

Experimental National Bureau of Standards SEAC-type computing circuits, some of considerable complexity, have been constructed and operated successfully at clock frequencies as high as 30 megacycles. No changes need to be made in the BTDL parameters when clock frequencies are altered. A description of these circuits is properly the subject of another paper.

Conclusions

A new type of diode switching circuit has been presented which has several unique advantages over conventional diode switching circuits at high speeds. The normally detrimental back transients associated with semiconductor diodes are here utilized as an essential feature of these circuits, which therefore have been named back-transient diode logic. Quiescent power dissipation is not only much reduced but at the same time has been freed from its dependence upon switching speed. Low-priced and readily available diodes may be used. It is felt that BTDL offers a new approach to designers of high-speed switching circuits worthy of their serious consideration.

References

1. ELECTRONS AND HOLES IN SEMICONDUCTORS (book), **W. Shockley.** D. Van Nostrand Company, Inc., Princeton, N. J., 1950.

2. ON SOME TRANSIENTS IN THE PULSE RESPONSE OF POINT-CONTACT GERMANIUM DIODES, **M. C. Waltz.** *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 40, Nov. 1952, pp. 1483–87.

3. RECOVERY CURRENTS IN GERMANIUM PN JUNCTION DIODES, **R. G. Shulman, M. E. McMahon.** *Journal of Applied Physics*, New York, N. Y., vol. 24, Oct. 1953, pp. 1267–72.

4. RECOVERY TIME MEASUREMENTS ON POINT-CONTACT GERMANIUM DIODES, **T. E. Firle, M. E.**

McMahon, J. F. Roach. *Proceedings*, Institute of Radio Engineers, vol. 43, May 1955, pp. 603–07.

5. Back-Transients in Semiconductor Diodes, C. G. Dorn. *NAVORD Report 4573*, U. S. Naval Ordnance, Washington, D. C., Apr. 1956.

6. Forward Transients in Point Contact Diodes, C. G. Dorn. *Transactions*, Professional Group on Electron Devices, Institute of Radio Engineers, vol. PGED-3, no. 3, July 1956, pp. 153–56.

7. Very Narrow Base Diodes, R. H. Rediker, D. W. Sawyer. *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 45, July 1957, pp. 944–53.

8. Some Silicon Junction Diode Recovery Phenomena, T. E. Firle. *WESCON Convention Record*, Institute of Radio Engineers, pt. 3, Aug. 1957.

9. Digital Switching Circuits, C. H. Page. *Electronics*, New York, N. Y., vol. 21, Sept. 1948, pp. 110–18.

10. Rectifier Networks for Multiposition Switching, D. R. Brown, N. Rochester. *Proceedings*, Institute of Radio Engineers, vol. 37, Feb. 1949, pp. 139–47.

11. Diode Coincidence and Mixing Circuits in Digital Computers, T. C. Chen. *Ibid.*, vol. 38, May 1950, pp. 511–14.

12. Digital Computer Components and Circuits (book), R. K. Richards. D. Van Nostrand Company, Inc., 1957.

13. Pulse and Digital Circuits (book), Jacob Millman, Herbert Taub. McGraw-Hill Book Company, Inc., New York, N. Y., 1956.

14. Computer Development (SEAC and DYSEAC) at the National Bureau of Standards. *Circular 551*, National Bureau of Standards, Washington, D. C., Jan. 1955.

A Nondestructive Breakdown Phenomenon in Selenium Rectifiers

A. C. ENGLISH
ASSOCIATE MEMBER AIEE

W. H. TOBIN
NONMEMBER AIEE

UNDER CONDITIONS of excessive current or voltage overload a rectifier will exhibit some form of breakdown. Under practical conditions, breakdown is normally destructive although, with proper safeguards, it may be controlled. A voltage-induced electric breakdown is the type most commonly cited in theory, while in practice a considerable contribution from thermal effects is observed. Selenium rectifiers frequently exhibit the property of recovering from an overvoltage breakdown. The symptoms in this case range from the appearance of minor sparking and fumes of selenium to a very generous arc blowing a cavity in one of the cells. Two physical mechanisms of recovery are apparent; in one, small areas of separation occur between the selenium and the counter-electrode, without necessarily rupturing the latter; in the other, greater burnouts involve craters reaching down to the baseplate, with walls of insulating vitreous selenium, and all counter-electrode removed from the immediate vicinity.

It is possible to observe another kind of failure which, for want of a better name, is labeled "stabilized breakdown" (SB), a nondestructive, essentially reversible phenomenon involving a thermal "runaway" to a limiting stabilization point.[1]

Experimental

Primary Phenomenon

A small size, e.g., $1^1/_2$ inches square, selenium rectifier is chosen, with a counter-electrode, such as cadmium, having a melting point higher than that of selenium, and connected to a normal a-c source and resistive load. Without imposing excessive voltages on the cell an overload condition is induced by increasing the forward current and, or, by raising the ambient temperature. While some cells show remarkable stability even above 200 C (degrees centigrade), all cells succumb to some combination of current and heat and enter a thermal "runaway" which continues until, nearing the melting point of selenium, it is halted by a stabilizing process. The events are best described in terms of typical instrumentation, whereby the current-voltage characteristics of the rectifier are displayed on an oscilloscope, and measurements are made of cell temperature (superficial), ambient temperature, and the output d-c average through the load.

Fig. 1 shows a set of concurrent events for a particular selenium cell under conditions where only the ambient temperature is varied. As the temperature rises the reverse current increases until, above 160 C, its effect appears as a reduction in the d-c output, while the cell temperature rise shows an acceleration because of increased heat generation. The cell is by now unstable and "runs away" until two distinguishable events interfere. At the first, a peak in reverse current is reached and a rapid fall to a lower value occurs. The output d-c current rises somewhat in response and the cell temperature continues its unabated climb. A momentary rise in reverse current is visible before the onset of the second event, in which, simultaneously, the output d-c current falls precipitously to a low value, the reverse current and forward current fall sharply, and the cell temperature stabilizes at a thermocouple reading between 210 and 220 C depending on the heat transfer conditions. The voltage-current patterns on the oscilloscope, Fig. 2, show that at 180 C (temperatures meas-

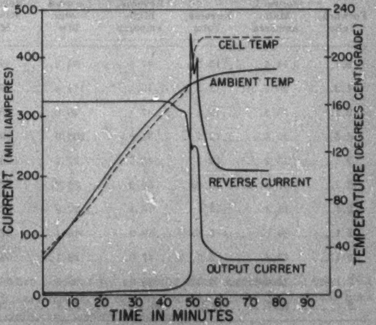


Fig. 1. Thermal runaway and SB of selenium rectifier. The cell is under a-c conditions and curves are drawn qualitatively from meter and oscilloscope data. Cell temperature is thermocouple reading on aluminum baseplate

---

Paper 58-821, recommended by the AIEE Solid State Devices Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting and Air Transportation Conference, Buffalo, N. Y., June 22–27, 1958. Manuscript submitted April 7, 1958; made available for printing October 2, 1959.

A. C. English is with the University of California, Berkeley, Calif.; and W. H. Tobin is with the General Electric Company, Lynn, Mass.

The authors wish to acknowledge the collaboration of P. J. Colleran and A. Just in the presentation of the earlier work.[1]

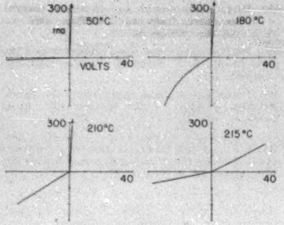


Fig. 2. Oscilloscope current-voltage patterns of selenium rectifier at several temperatures

210 C—reverse quadrant shows SB changes
215 C—whole cell is involved
Temperatures are thermocouple readings on the aluminum baseplate

ured on the aluminum plate) the reverse current is high while the forward characteristic is very little changed. At 210 C the reverse current has suddenly diminished. At 215 C the whole pattern swings uniformly, in the direction of increased resistance in both forward and reverse direction.

In the state now existing the cell shows a remarkable response to changes in imposed conditions. Within the limits of thermocouple measurements under variable heat transfer rates, the cell temperature is remarkably constant (about 1 C with only moderate precautions in technique) even when the ambient temperature is varied 50 C, the series load resistance varied widely, or the input voltage increased two or three times. On the oscilloscope, the current-voltage trace can be made to swing at will in response to the various stimuli. A change which would normally raise the temperature of the cell by any of these methods appears to result merely in an increase of over-all cell resistance. This effectively reduces the electric power being dissipated in the cell and so compensates with negligible change in cell temperature.

In anticipation of more detailed discussion, the phenomenon described would be quite plausible if the temperature were held constant because of a phase change, such as melting. The obvious material in this experiment, in view of the temperature, is selenium. A number of experiments help to clarify the details.

TEMPERATURE STABILIZATION

An attempt was made to determine with some accuracy the temperature at which the cell stabilized. The main problem lies in the fact that the region of highest temperature, presumably the rectifying junction, is inaccessible to measurement and that it is too difficult, because of destructive breakdown, to experiment with SB under equilibrium temperature conditions. The best compromise, a slowly rising temperature in well-insulated conditions, is limited by the necessity of clearly establishing that the cell actually experiences SB when the voltage is of necessity reduced.

A thermocouple was calibrated under our experimental conditions against the melting point of pure tin (accepted value 231.9 C) and then used to measure the selenium cell temperature, in SB, with the cell well insulated in packed asbestos and the temperature slowly rising. Because of the restricted heat transfer it was difficult to prevent destruction of the cell in a short time, although the electrical conditions were controlled. In one experiment with the thermocouple on the aluminum base-plate, the last 4 minutes of the run before a short circuit occurred showed a temperature change from 218.9–219.3 C. With the thermocouple on the cadmium counter-electrode the last 8 minutes showed a temperature change from 219.6–220.7 C. All factors considered, these temperatures as listed are probably not reliable within 2 C as a measure of the rectifying junction temperature.

The essential constancy of the temperature of a cell under SB was established under a wide range of electrical conditions, which include a-c, d-c, both forward and reverse, and a variety of square waveforms of 60 cps (cycles per second), with independent reverse and forward voltage values. One interesting confirmation of a constant temperature condition is found in the fact (considered later) that in constant ambient temperature conditions, wide variations in the imposed electrical conditions do not significantly change the amount of energy being liberated within the rectifier. The sensitive temperature measurements made while studying transient effects show cell temperature variations of a few tenths of a degree C for steady-state SB conditions.

ELECTRICAL PHENOMENA

The early experiments on SB were performed under typical a-c conditions with pure resistance or resistance- and capacitance-type loads. Oscillographic analysis was sufficient for a semiquantitative description of the events taking place. For more detailed analysis, a variety of conditions were investigated, in-

Table I. Current, Voltage, Resistance, and Power Dissipation of Rectifiers in Various States Within the SB Region

1	2	3	4	5	6	7	8	9	10	11	12	13
Forward, Volts	Forward, Milliamperes	Reverse, Volts	Reverse, Milliamperes	Forward Power, Mw	Reverse Power, Mw	Total Power, Mw	Forward Resistance, Ohms	Reverse Resistance, Ohms	Forward Removed Parallel Resistance, Ohms	Reverse Removed Parallel Resistance, Ohms	Forward Added Series Resistance, Ohms	Reverse Added Series Resistance, Ohms
2.1	14.4	12.5	47.5	30.2	595	625	146	263				
									2,020	1,670	12	49
4.2	26.7	12.5	40.2	111.8	502	614	158	312				
									2,150	1,755	11	47
2.1	14.3	12.5	47.1	30.0	590	620	147	265				
									618	790	46	131
6.3	32.6	12.5	31.6	205.0	395	600	193	396				
									900	840	34	127
2.1	13.2	12.5	46.5	27.7	581	609	159	269				
									490	611	39	80
2.1	17.5	10.45	55.2	36.7	578	615	120	189				
									499	625	38	82
2.1	13.3	12.5	46.1	27.9	578	606	158	271				
									218	218	67	151
2.1	23.1	8.35	69.2	48.4	579	627	91	120				
									218	218	65	146
2.1	13.4	12.5	47.0	28.1	589	617	156	266				
									114	97	90	195
2.1	31.6	6.3	88.9	66.3	560	626	66	71				
									116	96	87	200
2.1	13.7	12.5	46.1	28.8	578	607	153	271				

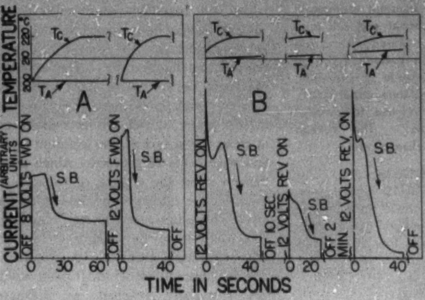


Fig. 3. SB induced by forward and reverse direct current

A—SB caused by forward current alone. For this experiment, voltages, though not currents, are much higher than usual

B—SB caused by reverse current alone. Ambient temperature T_a, cell temperature T_c and currents are taken from strip-chart records

cluding direct current, so as to separate completely the forward and reverse characteristics, and square-wave input derived from batteries and a 60-cps vibrator-chopper to evade some of the complexities of analyzing nonlinear characteristics and sine waves.

One experiment using the square-wave input was designed 1. to show the constancy of electric energy dissipated in the cell, and 2. to investigate the resistance changes in the cell. Under constant ambient temperature conditions and with the cell in the SB condition, changes were made in the forward or reverse voltage on the cell and measurements made of forward and reverse voltage and current. These data are recorded in the first four columns of Table I. Columns 5 and 6 show the power dissipated in the cell in the forward and reverse directions, and column 7 the total power dissipated. It will be observed that the total power column shows an average deviation from the mean of only 1% (even excluding repeated identical settings).

The same data were further analyzed to determine in what way the melting process added resistance to the cell. Of the four obvious cases tested it was found that a change in the SB condition had the effect of removing or adding a resistance in parallel with the cell rather than one in series, or than a nonlinear resistance in series or parallel, where the nonlinearity is matched to the cell characteristic at the time.

In Table I, columns 8 and 9 give the forward and reverse resistance of the cell at the various SB settings. Columns 10 and 11 may be compared for equality. They represent the "removed" parallel resistance when the cell at a higher power level is compared with its state at a lower power level and the two columns give, respectively, the values for forward and reverse measurements. Columns 12 and 13 give similar figures for the "added" series resistance, and comparison shows this case to be a poorer choice. It was verified that both voltage and current in both directions were essentially square waves and that no obvious changes in cell characteristics occurred during the 1/120th second of the cycle.

Fig. 3(A) shows the behavior of the cell when SB is induced by forward direct current alone. The chart record shows the cell entering the SB condition soon after 8 volts forward current is applied, recovering during an off period and repeating the process at 12 volts forward current. It will be seen that before SB a higher voltage gives higher current, afterwards, a lower current.

The behavior for reverse direct current shown in Fig. 3(B) is somewhat more complex. Usually the reverse current rises to a high value just before SB and in most cases then passes through a sharp fall and small rise before dropping to a steady low value. This "second peak" of current is most easily seen when the cell moves rapidly into SB and can be seen in all the circuit conditions when reverse voltage is present if the energy input is high enough.

An almost trivial experiment was devised to demonstrate that some of the phenomena of SB may be observed when only a part of the selenium rectifying area is in SB as might be the case when there are local hot-spots on a cell. Under a-c conditions, two cells were connected in parallel, with one of the cells in an oven. It was then possible to subject the one cell to SB while the other functioned normally.

Since the reverse current of such a combination is essentially dominated by the hotter cell the events recorded in reverse are quite typical of SB. Thus the fact that the reverse SB phenomena are observed is no guarantee that the whole of the cell surface is participating. Likewise, in the forward direction, the combination shows mainly the characteristics of the cool cell even though a part of the rectifying area had entered SB. Thus, it is possible to observe reverse SB phenomena without, or before, observing clear-cut changes in the forward direction.

Fig. 4 is a strip-chart recording of events while subjecting a rectifier to 5 cycles of entering and leaving the SB condition, the controlling variable being the ambient temperature T_1. The power source in this case was chopped d-c, 2 volts in the forward direction, 12 volts in reverse. The chart is of interest as a record of some complex events, not otherwise easily described. The inter-

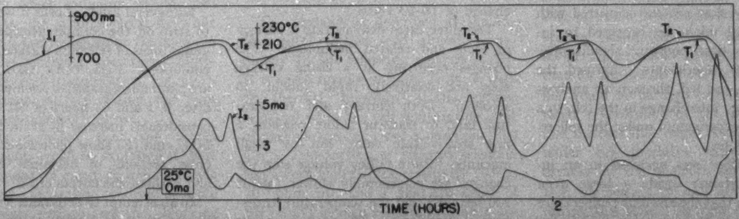


Fig. 4. **Rectifier subjected to 5 cycles in and out of SB. The current is a square wave obtained by chopping direct current at 60 cps. The forward current I_1 is at 2 volts input, reverse current I_2 at 12 volts. The recorded cell temperature T_2 is sensitive to changes in the ambient temperature T_1 and is not an accurate measure of the rectifying junction temperature**

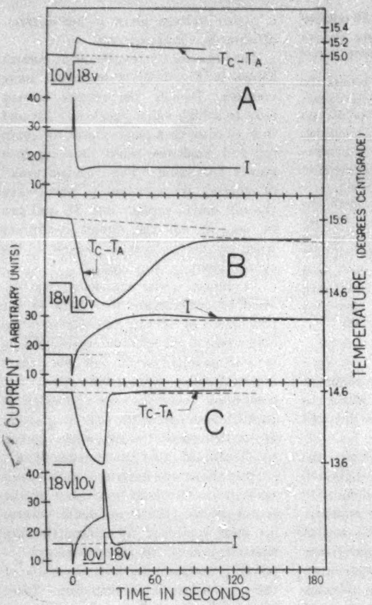


Fig. 5. Transient changes in reverse current and cell temperature caused by step changes in reverse d-c voltage on a rectifier in SB condition. Cell temperature, about 220 C, is given relative to ambient temperature. The voltage changes are

A—From 10–18 volts
B—From 18–10 volts
C—From 18–10 volts, and, at the temperature minimum for the transient, from 10–18 volts

relationship between the various curves is of primary interest. The peaks in the reverse current show a smooth trend with time, the forward current curve is somewhat less simple.

TRANSIENT PHENOMENA

The concept that SB depends upon a part of the selenium existing in the molten state suggests that in making a change of input voltage a part of the selenium should either freeze or melt. The time constant of the change in current observed after a step change in the voltage should then be related to the speed with which the two processes, freezing and melting, can occur. In selenium, crystallization is a relatively slow process compared with melting and might be expected to increase the time constant significantly. Preliminary experiments confirmed the predictions and led, through an analysis of the energy interchange in the cell, to a variety of experiments under the following conditions.

The rectifier was enclosed in an insulated container inside a large oven arranged so that the ambient temperature changes could be kept below 0.05 C per minute and with only long-period drift. Two opposing thermocouple junctions measured the temperature difference (about 20 C) between the air in the container and the rectifier. The resulting voltage was measured directly, or was balanced almost to zero and then amplified so that the recorder chart sensitivity became 0.07 C per cm (centimeter). (Noise level was about 0.3 cm.) The rectifier was subjected to step changes of reverse voltage in the range 10–18 volts and the current recorded. Fig. 5 shows the results of three switching operations: Fig. 5(A) from 10–18 volts, Fig. 5(B) from 18–10 volts, and finally Fig. 5(C) where the voltage was switched from 10–18 volts at the moment when the cell temperature had reached a minimum during a transient resulting from a change of 18–10 volts.

The three cases demonstrate a variety of facts and related deductions. The changes occurring at a rising voltage step are relatively rapid (about 15 seconds). Both current and temperature tend to transcur before reaching a new steady-state value, but by small amounts. For a falling voltage step the changes are considerably slower (about 100 seconds) and the temperature transcursion very large. The positive energy increment which the cell receives during the transient in Fig. 5(A) is considerably smaller than the negative increment of (B) which fact, together with other details in the events, may be explained by changes in heat transfer to the ambient during temperature fluctuations. The essential difference between Figs. 5(A) and 5(B) may be correlated with a melting process, in the former, driven by heat produced in, or on the surface of, the selenium itself, and with a freezing process which must depend upon the cooling of the cell to produce the supercooling which is the driving force in the latter case. A superficial analysis of the recordings for Fig. 5(B) indicates that at the temperature minimum (at 30 seconds) the freezing has essentially reached its ultimate steady-state condition, since the current also is very close to its final value. However, since the cell temperature is 1 C lower than it final value, crystallization continues to about the 60-second mark, with a coincident current overshoot which provides the increment of energy required to bring up the cell temperature and to remelt the excess of crystals. A similar argument applies to the much smaller temperature overshoot in Fig. 5(A) except that if the "superheating" of a crystal above its melting point is not permitted than the higher temperature requires additional explanation. Two possibilities seem to be that a semiconductor like cadmium selenide on the selenium surface may be a source of heat and can reach higher temperatures, or that selenium grains, of higher purity than the melt, may reach a higher temperature without passing their melting point and would not dissolve in the melt faster than the impurities in the melt diffuse to the grain surface.

MICROSCOPE STUDIES

A number of investigations were made to determine whether a momentary or prolonged experience of SB resulted in identifiable physical changes in the cell structure. It was natural to expect some obvious effects because of the high temperature.

On several kinds of cells it was possible to strip off the counter-electrode and examine areas of selenium surface in various conditions. Figs. 6(A), (B), and (C), are photomicrographs of selenium surfaces after 0, 1 and 63 hours of SB showing a considerable increase in grain size. Fig. 6(D), and (E) show photomicrographs of cross sections of selenium cells with and without the effects of SB.

In an attempt to determine whether the degree of grain growth at the selenium surface could be explained completely by the temperature of the cell during SB, a selenium cell was prepared in which a

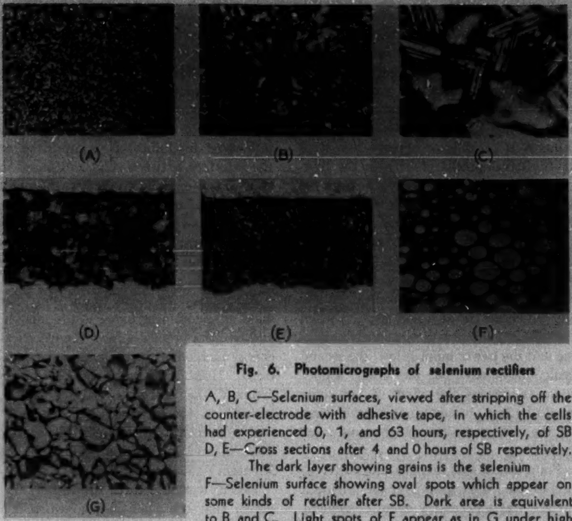


Fig. 6. Photomicrographs of selenium rectifiers

A, B, C—Selenium surfaces, viewed after stripping off the counter-electrode with adhesive tape, in which the cells had experienced 0, 1, and 63 hours, respectively, of SB
D, E—Cross sections after 4 and 0 hours of SB respectively. The dark layer showing grains is the selenium
F—Selenium surface showing oval spots which appear on some kinds of rectifier after SB. Dark area is equivalent to B and C. Light spots of F appear as in G under high magnification where the crystal grains look fire-polished
All except F are magnified 250×, F is 2.5×

small area was isolated electrically by slicing down to the aluminum with a razor blade. After SB the grain growth of the selenium surface which had not experienced the electrical conditions of SB was only about one third that of the rest of the cell. It seems unnecessary to search for more complex explanations than the temperature gradients which must exist between the rectifying junction under SB and any other parts of the cell.

On some cells, stripping the counter-electrode after SB uncovered selenium surfaces on which varying densities of irregular oval spots could be seen. A "negative" effect, Fig. 6(F), could be seen on the inner surface of the counter-electrode. Microscopic studies showed that the oval spots on the selenium were areas over which cadmium selenide had adhered to the counter-electrode while elsewhere it had remained on the selenium during the stripping operation. Moreover, under high magnification, Fig. 6(G), the selenium surface within the spot had a ' fire-polished" aspect.

It was surmised that the aspect of these spots could be explained if there was a sequence of events in which at some early stage a separation occurred between the selenium surface and the "blocking layer" and subsequently this free selenium surface was subjected to high enough temperatures that vaporization of the selenium would fire-polish the crystal grains. This concept was supported by experiments in which the characteristics of Fig. 6(G) were duplicated, first on a stripped selenium surface, open to the air during an SB run on adjacent rectifier area; and second on a similar partly stripped cell in which a piece of mica was laid over the open selenium to simulate, as for the oval spots, the presence of the counter-electrode. In both these cases the selenium appearance obtained was indistinguishable from that in the oval spots, and a deposit of selenium crystals on the mica sheet gave ample evidence of a mechanism involving evaporation of selenium.

Discussion

The phenomena described in the experimental section lead almost immediately to the following analysis of SB. It is based mainly on three facts: 1. that the thermal "runaway" of the rectifier is halted when the cell temperature is at or near the melting point of selenium (220 C), 2. that molten or vitreous selenium has a very high electric resistivity (10^5 to 10^{12} ohm-cm), and 3. that the onset of SB results in high electric resistance appearing suddenly in the cell in a manner peculiarly sensitive to imposed conditions which would normally produce a temperature change. The essential result is that once the cell enters SB a controlling mechanism appears which maintains the cell at almost a constant temperature over a very wide range of electrical or ambient temperature conditions. The "thermostat" is quite simple in function—for any increase in ambient temperature or forward or reverse voltage, the cell resistance increases, and the power dissipated in the cell E^2/R decreases enough to compensate, and vice versa.

It must first be recognized that a high proportion of the electric power dissipated in the cell appears in a very thin layer, the rectifying junction, and physical changes in this layer could completely account for all the observed phenomena. It is not necessary to maintain the junction, and the heat generated, within the selenium. A second semiconductor, such as a metal selenide, which may be naturally[2] or artificially present on the surface, so long as its melting point is higher than that of selenium, could function as the hottest layer. The minimum requirement seems to be that in the region of highest temperature, at the selenium surface, the path of current flow should be obstructed by molten selenium. On the question of how much selenium melts, it might be argued that an extremely thin layer (say 0.1–1 micron) would be enough to account for the increased resistance and that no more than this could be formed without violating the thermostatic principles. Since the selenium is a polycrystalline mass it is unlikely that the current flow or the thickness of the rectifying junction is uniform, so that the regions which melt may be of peculiar shape, as for example thin filaments of grain boundaries. It is possible that the crystal grains themselves are of high resistivity and that the doping impurity is segregated to the grain boundaries, which latter may be the primary current path, either of themselves or through the influence of surface charges on the grains. Whether this is so or not, grain boundaries are probably of prime importance in the melting process postulated since their melting temperatures would be somewhat lower than for the grains. Any extensive melting would be hard to correlate with the rapid reversibility of the phenomenon, and the fact that many hours of running under SB produces extremely large crystals on the selenium surface implies that the surface crystals are not molten during this time. In view of the claim by Burke[3] (brought to the author's attention by U. Martius) that melting of grain boundaries prevents grain growth, grain boundary melting below the uppermost layer of grains is not indicated.

If the controlling mechanism depends upon a very small amount of molten selenium at the rectifying junction (and the amount need not vary much to cover the experimental range of conditions), the essentially constant temperature of the cell can be easily understood. This is best observed with heat-transfer conditions from the cell to its environment held constant when, as described previously, it is found that the power dissipated within the cell is controlled at a constant value in spite of changes in the forward or reverse voltage. The change in the amount of molten selenium necessary to adjust the cell resistance would not be readily detectable since the structure and temperature distribution in the cell are not significantly modified.

From the treatment of data given in Table I it would appear that, if one regards the rectifier as an aggregate of parellel current carriers, the increased melting of selenium has the effect of removing some of the parallel conduction from the system. The picture would be a little more pleasing if the conductance removed had the identical nonlinear characteristic of the cell as a whole at the moment of change, but the evidence at present favors, if anything, a conductance which is about the same for forward and reverse. This may mean that in the general inhomogeneity of the rectifying junction the regions that will be involved in a change of SB conditions are those which are least rectifying in character, i.e., those which have deteriorated the most in rectifying ability.

In addition to inhomogeneities on a microscopic scale there are certainly larger scale irregularities, commonly thought of as "hot spots." It is convenient to explain some of the secondary phenomena of SB with the aid of hot spots. The presence of two peaks in the reverse current as a cell moves into SB may be explained as follows. During the thermal runaway the hot spots, being highest in temperature, are the first areas to reach the SB condition, and are therefore the first areas to change from contributing largely to reverse current to passing much smaller currents. Thus the large, first peak seen in the reverse current would represent the most "leaky" areas of a cell reaching the SB condition. The second peak, which is quite variable in form and size, seems to represent first the continued increase of reverse current through areas which are still rising in temperature, and finally the fall in reverse current when the total rectifying area enters SB. This idea is made more plausible by the fact that a distinct and rapid change in the forward characteristic occurs only at the second peak. Thus it is possible to observe two steps in the development of SB; in the first, only a part of the cell is involved and the main changes are in the reverse characteristics, in the second, the whole cell participates and forward and reverse changes are simultaneously observed. An experiment has been described which represents part of this process. When the cell enters SB at a slow rate, as when the ambient temperature is slowly changing and the power input is low, the second peak may not be observed. Under these conditions the effect of hot spots would be less obvious since there is time for the cell temperature to become more uniform.

For a time it was thought that the creation of the "oval spots" on the selenium surface was a necessary event coincident with the first main peak in the reverse current, i.e., hot spots were irreversibly being removed by open circuit. However, it appears that the oval spots are not always present, and whether or not they represent an irreversible change, the observed electrical phenomena are always about the same, and essentially reversible; which is, perhaps, to say that there are various kinds of hot spots always available, as would be expected.

The experiments analyzing the freezing and melting of selenium may be discussed from several viewpoints. When the voltage input to the cell is suddenly reduced, the cell does not adjust instantaneously. Nor does the freezing process deliver heat to the cell rapidly enough to maintain constant temperature until the correction of cell resistance returns the power dissipation to the SB level. The change in temperature experienced by the cell would seem to violate the steady state requirements of an equilibrium temperature controlled by a phase change until it is realized that the cell continues to exist in SB by virtue of the presence of insulating selenium rather than through the temperature *per se.* Transient conditions can result in temporary supercooling of the melt without changing the electrical properties, upon which the supercooling drives the system back to equilibrium at a rate controlled by the crystallization velocity.

It is possible that the major part of the difference in velocity between crystallizing and melting is caused by the difference in times required to orient and disorient selenium chains which exist in both the liquid and the solid state. However, both of the processes are probably influenced by diffusion of impurities at the interface between solid and liquid in view of two experimental facts: the slight overshoot of temperature in the melting transient and the fact that the steady-state temperatures before and after a voltage step differ by tenths of a degree centigrade; see Fig. 5(A) for example

Attempts were made to calculate the mass of selenium involved in the melting and freezing processes from the electrical energy increments in the transient conditions. However, it was never possible to estimate the energy involved in heat transfer to the environment sufficiently well to identify convincingly the energy used in melting a few milligrams of selenium (0.065 joule per milligram).

Recent work by one of the authors has shown that the phenomenon which has been described in this paper provides an almost complete explanation of creep and forming effects in the selenium rectifier.

Conclusions

A unique phenomenon in selenium rectifiers is described and analyzed. The rectifier is subjected to overload conditions and after a thermal runaway suddenly stabilizes near the melting temperature of selenium. Under the stabilized condition it is hypothesized that the rectifier is held at almost constant temperature under a wide range of electrical and ambient conditions by a sensitive mechanism which compensates for a change imposed on the cell by melting or freezing selenium. This adjusts the cell resistance and thereby the electric power dissipation in the cell.

References

1. A NON-DESTRUCTIVE BREAKDOWN PHENOMENON IN SELENIUM RECTIFIERS (Abstract), **A. C. English, P. J. Colleran, A. Just, W. H. Tobin.** *Journal of the Electrochemical Society*, Baltimore Md., vol. 101, no. 4, Apr. 1954, p. 95c ff.

2. ON THE BARRIER LAYER (FORMED) BY THE CONTACT BETWEEN THE SELENIUM SEMICONDUCTOR AND METAL. II. ON THE SELENIDE FORMATION IN THE BARRIER LAYER, **M. Tomura.** *Bulletin*, Chemical Society of Japan, Tokyo, Japan, 22, 1949, pp. 82–91.

3. SOME FACTORS AFFECTING THE RATE OF GRAIN GROWTH IN METALS, **J. E. Burke.** *Transactions*, American Institute of Mining and Metallurgical Engineers, New York, N. Y., vol. 180, 1949, pp. 73–91.

The Effect of Source and Sink Thermal Resistance on Thermoelectric Generator Performance

P. E. GRAY
NONMEMBER AIEE

A STATIC ANALYSIS of the thermoelectric generator is necessary in order to specify the electrical loading condition for maximum energy conversion efficiency and to determine the corresponding efficiency. Ioffe has done this for the simple case of a single *p-n* junction thermocouple, and his analysis has been extended to include the effect of temperature dependent parameters by Borrego, Lyden, and Blair.[1,2] Harman has considered a cascade generator and has shown how its design can be optimized.[3]

All previous writers on the subject have assumed that the temperatures of the thermal source and sink are fixed. It is also of interest to investigate the static analysis of a thermoelectric generator for the more realistic case in which the source and sink temperatures are not fixed but depend upon the heat flows from the source and to the sink. This case of nonideal thermal sources becomes of greater importance as the conversion efficiency of the device is increased, since the electrical loading will then have a greater effect on the heat flows and thus on the temperatures at the ends of the couple.

The purpose of this paper is to analyze the static performance of a thermoelectric generator which operates between a thermal source and a thermal sink having nonzero thermal resistances. In particular, the dependence of junction temperatures and heat flows on electrical loading is determined. Also, the effect upon the condition for optimum loading and upon the optimum efficiency is assessed. It is shown that the corrections produced by this analysis are not negligible in several numerical examples.

The Device Model

The model which has been studied is shown in Fig. 1. T_h and T_c represent the fixed zero-heat-flow temperatures of the source and sink respectively. T_l and T_0 represent the corresponding temperature of the hot and cold junctions. P_l and P_0 are the heat flows out of the source and sink, and R_h and R_c are the thermal resistances of the source and sink respectively. It was assumed that the temperatures and heat flows can be related by the following equations

$$T_l = T_h - R_h P_l \qquad \text{(1A)}$$

$$T_0 = T_c - R_c P_0 \qquad \text{(1B)}$$

These linear relationships between junction temperatures and the power flows may be regarded as the first terms of power series expansions of the junction temperatures. They are assumed to be adequate approximations for the case being considered.

It is also assumed that the thermal conductivities of the two legs of the couple k_1 and k_2, the electrical resistivities ρ_1 and ρ_2, and the Thomson coefficients, τ_1 and τ_2, are not functions of temperature. The Peltier coefficients of the two junctions are $\pi_l = \alpha_l T_l$ and $\pi_0 = \alpha_0 T_0$ where α_l and α_0 are the Seebeck coefficients of the two junctions, which will be functions of temperature. The two legs of the couple are of uniform cross-sectional areas A_1 and A_2 and are both of length l. The electrical terminal voltage and current are E and I respectively.

Determination of Junction Temperatures and Heat Flows

It is first necessary to determine the heat flows P_l and P_0, as functions of the junction temperatures T_l and T_0, and of the electric current I. This problem has been solved for the one-dimensional case in which the total Thomson heat is small compared to the conduction heat.[1] This assumption with the results thus obtained will be used here in the following forms

$$P_l = \alpha_l T_l I + \left(\frac{k_1 A_1}{l} + \frac{k_2 A_2}{l}\right)(T_l - T_0) + \frac{1}{2} I(\tau_1 - \tau_2)(T_l - T_0) - \frac{1}{2} I^2 \left(\frac{\rho_1 l}{A_1} + \frac{\rho_2 l}{A_2}\right) \qquad \text{(2A)}$$

$$P_0 = -\alpha_0 T_0 I - \left(\frac{k_1 A_1}{l} + \frac{k_2 A_2}{l}\right)(T_l - T_0) + \frac{1}{2} I(\tau_1 - \tau_2)(T_l - T_0) - \frac{1}{2} I^2 \left(\frac{\rho_1 l}{A_1} + \frac{\rho_2 l}{A_2}\right) \qquad \text{(2B)}$$

Equations 2 can now be written in the following forms which will be more useful for our purposes

$$P_l = g_{ll} T_l - g_{l0} T_0 - p_l \qquad \text{(3A)}$$

$$P_0 = -g_{0l} T_l + g_{00} T_0 - p_0 \qquad \text{(3B)}$$

in which the six coefficients are defined as

$$g_{ll} = \alpha_l I + \frac{1}{R_E} + \frac{1}{2} I(\tau_1 - \tau_2) \qquad \text{(4A)}$$

$$g_{l0} = \frac{1}{R_E} + \frac{1}{2} I(\tau_1 - \tau_2) \qquad \text{(4B)}$$

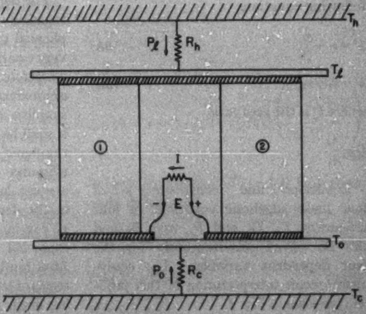


Fig. 1. Device model

Paper 60-18, recommended by the AIEE Solid State Devices Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., January 31–February 5, 1960. Manuscript submitted August 12, 1959; made available for printing October 8, 1959.

P. E. Gray is with the Massachusetts Institute of Technology, Cambridge, Mass.

The research reported in this paper has been sponsored by the Electronics Research Directorate of the Air Force Cambridge Research Center, Air Research and Development Command, under contract *AF33(616)-3984*. Although this paper is based on work sponsored by the U. S. Air Force it has not been approved or disapproved by that agency.

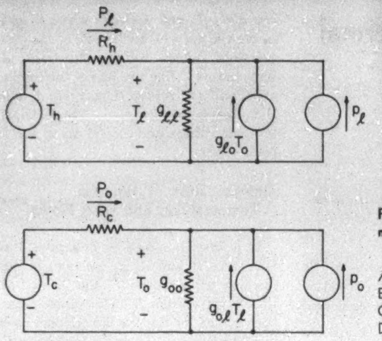


Fig. 2 (left). Equivalent circuit interpretation

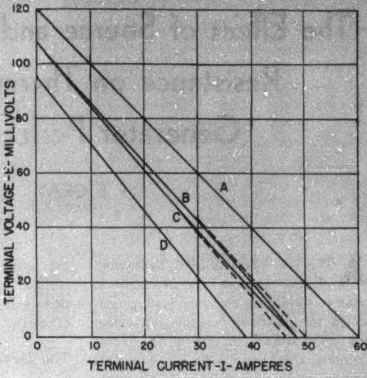


Fig. 3 (right). Terminal volt-ampere curves

A—x=0, y=0
B—x=0, y=0.1
C—x=0.1, y=0
D—x=0.1, y=0.1

$$p_l=\frac{1}{2}I^2r_E \qquad (4C)$$

and

$$g_{00}=-\alpha_0 I+\frac{1}{R_E}-\frac{1}{2}I(\tau_1-\tau_2) \qquad (5A)$$

$$g_{0l}=\frac{1}{R_E}-\frac{1}{2}I(\tau_1-\tau_2) \qquad (5B)$$

$$p_0=\frac{1}{2}I^2r_E \qquad (5C)$$

Here two coefficients, R_E and r_E, have been defined for convenience, which are the total thermal resistance of the couple (two legs thermally in parallel) and the total electric resistance of the couple (two legs electrically in series) respectively

$$R_E=\frac{1}{\frac{k_1A_1}{l}+\frac{k_2A_2}{l}}=\frac{1}{\frac{k'A_1}{l}} \qquad (6)$$

$$r_E=\frac{\rho_1 l}{A_1}+\frac{\rho_2 l}{A_2}=\frac{\rho' l}{A_1} \qquad (7)$$

The latter forms of the foregoing equations express the thermal and electric resistances in terms of equivalent conductances and resistivities which will be useful later and which are defined by the following equations

$$\rho'=\rho_1+\frac{\rho_2}{Q} \qquad (8A)$$

$$k'=k_1+k_2Q \qquad (8B)$$

where Q is the area ratio

$$Q=\frac{A_2}{A_1} \qquad (9)$$

Equations 1 and 3 constitute a set of four linear algebraic equations in four unknowns T_l, T_0, and P_l, and P_0, and can be solved simultaneously for the four dependent variables. The equivalent circuit interpretation of this problem is shown in Fig. 2.

The solution of equations 1 and 3 for the heat flows is straightforward and the results are shown below:

$$P_l=\frac{P_{lf}(1-yR_Eg_{00})+P_{0f}(yR_Eg_{l0})}{(1+xR_Eg_{ll})(1+yR_Eg_{00})-(xR_Eg_{0l})(yR_Eg_{l0})} \qquad (10A)$$

$$P_0=\frac{P_{0f}(1-xR_Eg_{ll})+P_{lf}(xR_Eg_{0l})}{(1+xR_Eg_{ll})(1+yR_Eg_{00})-(xR_Eg_{0l})(yR_Eg_{l0})} \qquad (10B)$$

Some comment on the notation is necessary at this point. The subscript f on P_{lf} and P_{0f} in equations 10 means that these expressions are to be evaluated according to equations 2 with the understanding that the fixed temperatures T_h and T_c are to be used in place of T_l and T_0 respectively. Furthermore, the dimensionless parameters x and y represent the thermal resistances of the source and sink normalized to the thermal resistance of the couple which is defined by equation 6. That is,

$$x=R_h/R_E \qquad (11A)$$

$$y=R_c/R_E \qquad (11B)$$

The dependence upon x and y which is found in equations 10 is sufficiently complicated to make the subsequent analysis very complicated and difficult to interpret. Therefore it is desirable to make an approximation at this point which will simplify these expressions. Specifically it shall be assumed that x and y are small compared to unity. Now if the coefficients of x and y in equations 10 are not large compared to unity, terms in the denominator of order x^2, y^2, and xy can be dropped, and the linear terms expanded by the binomial theorem. After multiplying out and again dropping terms of second order or higher the following linear relations are obtained:

$$P_l\simeq P_{lf}(1-xR_Eg_{ll})+P_{0f}(yR_Eg_{l0}) \qquad (12A)$$

$$P_0\simeq P_{lf}(xR_Eg_{0l})+P_{0f}(1-yR_Eg_{00}) \qquad (12B)$$

In order to test the validity of this approximate expansion, second-order terms have been examined for the case in which the Peltier heat terms in equations 4 and 5 are 40% of the conduction terms and the Thomson heat terms are 10% of the conduction terms. These assumed values are large enough so that the estimate of error will be conservative for thermocouples with figures of merit as large as 3×10^{-3} even under short-circuit conditions. For the numbers chosen the error in P_0 and P_l made by neglecting terms in x^2, y^2, and xy in the expansion is less than 5% if $x=y=0.1$. Since this represents an upper bound on the error for practical thermocouples, it can be concluded that the approximations contained in equations 12 are not severe.

Equations 12 can now be substituted into equations 1 in order to determine the junction temperatures. To the first order in x and y these are

$$T_l=T_h-xR_EP_{lf} \qquad (13A)$$

$$T_0=T_c-yR_EP_{0f} \qquad (13B)$$

This completes the determination of the approximate junction temperatures and heat flows in the presence of source and sink thermal resistance. We have succeeded in expressing these quantities in such a way that all dependencies upon the electric current I are explicit. This was not true for equations 2 since the temperatures T_l and T_0 were at that point implicit functions of current. It should be noted at this point that the implicit dependence of α_l and α_0 on current through their temperature dependence has been ignored. This has been done because it is difficult to generalize about the temperature dependence of the Seebeck coefficient and also such a generalization would have led to equations relating heat flow and junction temperature which were nonlinear or transcendental. Furthermore, the error thus made is certainly small in a practical thermocouple.

Effect upon Terminal Volt-Ampere Characteristics

The effects of x and y, the normalized source and sink thermal resistances upon the terminal volt-ampere characteristics of the generator can readily be assessed by applying the principle of conservation of energy, which requires that

$$P_I+P_0=EI \tag{14}$$

When equations 12 are substituted into equation 14 a factor of I can be eliminated after considerable algebraic manipulation, and the following result is obtained for the terminal voltage

$$E=E_0-Ir+I^2s \tag{15}$$

E_0 the open-circuit voltage, r the coefficient of the linear term, and s the coefficient of the quadratic term are given by the following expressions

$$\begin{aligned}E_0=&(\alpha_l+\tau_1-\tau_2)[T_h-x(T_h-T_c)]-\\&(\alpha_0+\tau_1-\tau_2)[T_c+y(T_h-T_c)]\end{aligned} \tag{16}$$

$$\begin{aligned}r=r_E+&[xR_E\alpha_lT_h(\alpha_l+\tau_1-\tau_2)+\\&yR_E\alpha_0T_c(\alpha_0+\tau_1-\tau_2)]+\\&[xR_E(\alpha_l+\tau_1-\tau_2)-yR_E(\alpha_0+\tau_1-\tau_2)]\times\\&\left[\frac{1}{2}(\tau_1-\tau_2)(T_h-T_c)\right]\end{aligned} \tag{17}$$

$$s=\frac{1}{2}r_E[xR_E(\alpha_l+\tau_1-\tau_2)-yR_E(\alpha_0+\tau_1-\tau_2)] \tag{18}$$

As is clear from equations 14–18, the effects of thermal resistance on the volt-ampere characteristics are threefold:

1. The open-circuit voltage is reduced because of the temperature drop in the thermal resistances caused by conduction heat flow.
2. The apparent electric resistance is increased because of the additional temperature drop in the thermal resistances because of the Peltier heat flow (second term); and may be either increased or decreased for the same reason by Thomson heat flow (third term).
3. A nonlinear contribution to the terminal voltage is caused by the temperature drops in the thermal resistances because of Joulean heat flow.

In order to assess the relative importance of these three effects, numerical results have been obtained for a simple example in which the Thomson heat is neglected and the parameters shown in the Nomenclature apply.

When these numerical parameters are substituted into equations 16–18 the following expressions result:

$$\begin{aligned}&E_0=120(1-x-y)\text{ millivolts}\\&r=2+3.2x+1.6y\text{ milliohms}\\&s=13.3\times10^{-3}(x-y)\text{ millivolts/ampere}^2\end{aligned} \tag{19}$$

In Fig. 3 terminal-volt-ampere curves are shown for the generator previously described corresponding to the four choices of normalized thermal resistance indicated on the curves. For the numerical values shown $R_E=33$ K (degrees Kelvin)/watt. The dashed lines in Fig. 3 show the resulting curves for cases B and C if the quadratic term in equation 15 is ignored. It is clear from these curves that the quantitative effect of this term is not large, particularly at currents near the normal operating current of the generator. For example, if $I=20$ amperes, the error in voltage made by neglecting the quadratic term is less than 1% for both cases B and C. On the other hand, the quantitative effect of the increase in r is not negligible and would lead to an error in voltage of about 9% at $I=20$ amperes if it were neglected.

Calculation of Optimum Current

One would now like to determine the effect of the normalized thermal resistances upon the condition for optimum loading of the generator. This can be accomplished by determining the value of current, I, which maximizes the conversion efficiency of the generator. Defining the efficiency η as the electric power output per unit heat input one has

$$\eta=\frac{EI}{P_I}=\frac{E_0I-rI^2+sI^3}{P_{If}(1-xR_Eg_{Il})+P_{0f}(yR_Eg_{I0})} \tag{20}$$

One can maximize this efficiency by the usual procedure of setting $d\eta/dI$ equal to zero. The resulting expression for the optimum current is fifth order in I, and one must resort to numerical methods to extract a root. Therefore, it appears desirable at this point to make a further approximation which will reduce the order of the equation that must be solved to yield I_0, the optimum current.

The numerical example discussed in the previous section suggests one approximation which will reduce the order of the numerator of equation 20. It was pointed out then that for currents of the order of one-half the short-circuit current, the error in terminal voltage made by dropping the quadratic term was less than 1%. Therefore, the corresponding term will be neglected in the numerator of the efficiency expression.

Even with the foregoing approximation the problem is formidable and some simplification of the denominator is also necessary. When equations 2 and 4 are substituted in the denominator of equation 20 (using fixed temperatures T_h and T_c), and the resulting multiplications are carried out, it is found that the troublesome terms in I^2 and I^3 arise physically from the products of Peltier heat coefficients, Thomson heat coefficients, and Joulean heats. Accordingly a reasonable approximation would be to neglect terms such as these which in practical thermocouples are certainly small in comparison to products involving the conduction heat coefficient.

When these simplifying approximations are made the resulting equation for I_0 can be shown to be a quadratic since the coefficient of the cubic term reduces to zero. This equation for I_0 is

$$\begin{aligned}I_0^2\Big[-\frac{1}{2}&E_0r_E(1-x-y)+r\{\alpha_lT_h(1-x)-\\&\alpha_0T_0y-\alpha_l(T_h-T_c)x\}+\\&\frac{1}{2}(\tau_1-\tau_2)(T_h-T_c)(1-2x)\Big]+\\&I_0\left[2r\frac{(T_h-T_c)}{R_E}(1-x-y)\right]+\\&\left[-E_0\frac{(T_h-T_c)}{R_E}(1-x-y)\right]=0\end{aligned} \tag{21}$$

Equation 21 can be solved by formula for the optimum current which will maximize the efficiency. In order to reduce the algebraic complexity involved in doing this it will be assume that Thomson heat is negligible. In this case $\tau_1=\tau_2=0$ which implies that $\alpha_l=\alpha_0=\alpha$.

After making the approximations just outlined and dropping terms higher than first order in x and y the solution to equation 21 can be put in the following form by considerable algebraic manipulation

$$I_0=I_{00}[1-\epsilon(x,y)] \tag{22}$$

where

$$I_{00}=\frac{\alpha(T_h-T_c)}{r_E(M+1)} \tag{23}$$

and

$$\epsilon(x,y)=(x+y)+(xT_h+yT_c)Z\left(\frac{\frac{1}{2}ZT_h+M}{M(M+1)}\right) \tag{24}$$

In these equations Z, $\bar{T}$, and M have their usual definitions[1]

$$Z=\frac{\alpha^2R_E}{r_E}=\frac{\alpha^2}{\rho'k'} \tag{25A}$$

$$\bar{T}=(T_h+T_c)/2 \tag{25B}$$

$$M=\sqrt{1+Z\bar{T}} \tag{25C}$$

Table I. Comparison of Actual and Predicted Values of Optimum Current and Efficiency

Curve	x	y	Current at Maximum Efficiency, Amperes	I_0, Amperes	Maximum Efficiency, %	η_0, %
E	0.03	0.05	20.8	20.9	10.8	10.5
F	0.10	0.05	18.0	17.6	9.8	9.1

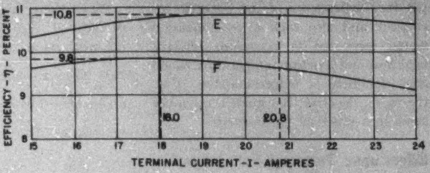


Fig. 4. Efficiency-current curves

E—$x=0.03$, $y=0.05$
F—0.10, $y=0.05$
Note supressed zeros

It is clear from equation 22 that the optimum load current is invariably reduced in the presence of thermal resistance. Equation 23 agrees with the results obtained by Ioffe for the zero-thermal resistance case although his analysis was put in terms of an optimum load resistance. Since the optimum load resistance is MR_E, and since $\alpha(T_h-T_c)$ is the open-circuit voltage, the current at the optimum point, which is the open-circuit voltage divided by the total resistance, is given by equation 23.

In the case of thermoelectric generators which use materials having a small figure of merit Z, the second term in equation 24 vanishes and the correction factor in equation 22 is simply $(1-x-y)$. To the first order in x and y this is the correction one would obtain if the junction temperatures T_l and T_0 were assumed to differ from the source and sink temperatures T_h and T_c only by the drop in the source and sink thermal resistances caused by the flow of conduction heat. The second term is therefore the consequence of a change in junction temperatures because of loading and becomes of greater importance as the figure of merit becomes larger.

The results presented in this section will be discussed in terms of two simple numerical example in the section on "Evaluation of Results." First we must consider the effect of thermal resistance on the optimum efficiency of the generator.

Calculation of Optimum Efficiency

The results of the preceding section can be used to evaluate the efficiency of the generator at the optimum current point. It is simply necessary to substitute equation 22 for the optimum current in equation 20 which expresses the efficiency as a function of current. Note that the cubic term in the numerator and certain quadratic and cubic terms in the denominator of equation 20 are neglected for the reasons discussed in the preceding section. The result of this substitution, correct to terms of first order in x and y, can be shown to be the following after considerable algebraic manipulation

$$\eta_0=\eta_{00}[1-\delta(x,y)] \tag{26}$$

$$\eta_{00}=\frac{(T_h-T_c)(M-1)}{T_h(M+T_c/T_h)} \tag{27}$$

$$\delta(x,y)=(x+y)\left[\frac{2M+1}{M}-2\frac{\dfrac{T_c}{T_h}+\dfrac{1}{M}}{\dfrac{T_c}{T_h}+M}\right]-$$
$$x\left[\frac{ZT_h}{2M}\,\frac{1-T_c^2/T_h^2}{M+T_c/T_h}\right]+$$
$$(xT_h+yT_c)Z\left[\frac{M-1}{M(M+T_c/T_h)}\right] \tag{28}$$

In these equations Z and M are as defined in equations 25(A) and (C). Note that the efficiency for the zero thermal resistance case, which is given by equation 27 agrees with results obtained by Ioffe. It is easy to show that equation 28 is non-negative for all non-negative values of x and y, so that the efficiency is invariably reduced in the presence of thermal resistance.

Again in the case of a thermoelectric generator fabricated of materials having a small figure of merit, Z the second and third terms in equation 28 become negligibly small. Furthermore, since M approaches unity the first term approaches $x+y$, and the correction factor in equation 26 is simply $(1-x-y)$. To the first order in x and y this is the correction one would obtain if the junction temperatures T_l and T_0 were assumed to differ from the source and sink temperatures T_h and T_c only by the drop in the source and sink thermal resistances caused by the flow of conduction heat. Therefore the second and third terms are consequences of the additional dependence of the junction temperatures upon the electrical loading through the Peltier heat and Joulean heat. They become of greater importance as the figure of merit is increased.

Note that the area ratio Q, which was defined by equation 9, has not yet been specified. To do so precisely would be difficult because not only is M a funtion of Q but also x and y, the normalized thermal resistances, are functions of Q for fixed value of R_h and R_c. Therefore, we will assume that the area ratio should be chosen in such a way as to maximize M, which will result in the maximum value of η_{00}. This is of course the same condition obtained by others[1,2] and requires the following value of Q.

$$Q=\left(\frac{\rho_2k_1}{\rho_1k_2}\right)^{1/2} \tag{29}$$

The results obtained in this section will be discussed for two simple numerical examples in the following section.

Evaluation of Results

The results of the preceding two sections will be checked by direct calculation of the efficiency versus current curves for two simple numerical examples in this section. Using the numerical parameters defined in the section on "Effect Upon Terminal Volt-Ampere Characteristics," curves of efficiency as a function of current have been computed directly from equation 20 without approximation for the two sets of values of x and y shown in Table I. The curves thus obtained are shown in Fig. 4 and the resulting values of optimum current and efficiency have been entered in Table I. Also the values of I_0 and η_0 predicted by equations 22 and 26 are shown in Table I. For comparison, I_{00} and η_{00} are 24.2 amperes and 12.1% respectively.

In general, the agreement between the actual values computed without approximation and the approximate values predicted by the results of the preceding two sections is quite good. It is clear from the values of I_{00} and η_{00} that large errors will be made if the junction temperatures are simply assumed equal to the internal source and sink temperatures. If on the other hand first-order corrections for the flow of conduction heat are made as previously discussed, the resulting values of I_0 and η_0 are 22.3 amperes and 11.1% respectively for case E and 20.6 amperes and 10.3% respectively for case F. In the case of the determination of optimum current, the results of this partial correction are still in error by amounts of about

7 and 14% for cases E and F respectively. In this case the complete first order correction presented in the section on "Calculation of Optimum Current" reduces the errors to less than 2% in both cases. In the case of the determination of optimum efficiency the errors are larger. The partial correction predicts an efficiency which is too large, whereas the complete first-order correction of the preceding section predicts an efficiency which is too small by about the same amount. The poorer agreement with regard to efficiency is apparently due to the neglect of higher order terms in the expansion in x and y. This conclusion is substantiated by the fact that the error is greater in case F than in case E and the second-order terms in x are larger in case F.

Since the efficiency-current curves of Fig. 4 are quite flat in the region near the maximum the question of the precise specification of an optimum value of load current may be somewhat academic. This is a consequence of the relatively low efficiency of practical thermoelectric generators and is present in the previously studied case. It is nevertheless desirable to know what constitutes optimum loading and approximately what optimum efficiency can be attained in any practical device and for this reason the preceding analysis is necessary.

Nomenclature

T_h, T_c = zero heat flow temperatures of heat source and heat sink
T_l, T_0 = temperatures of hot and cold junctions
P_l, P_0 = heat flows out of source and sink
R_h, R_c = thermal resistances of source and sink
k_1, k_2 = thermal conductivities of legs 1 and 2
$\rho_1 \rho_2$ = electrical resistivities of legs 1 and 2
τ_1, τ_2 = Thompson coefficients of legs 1 and 2
π_l, π_0 = Peltier coefficients of hot and cold junctions
α_l, α_0 = Seebeck coefficients at hot junction temperature and cold junction temperature
A_1, A_2 = cross-sectional areas of legs 1 and 2
l = length of active material
E = terminal voltage
I = terminal current
R_E = total thermal resistance of the couple
r_E = total electrical resistance of the couple
ρ' = equivalent electrical resistivity
k' = equivalent thermal conductivity
Q = ratio of the area of leg 2 to the area of leg 1
x, y = thermal resistances of source and sink normalized to the thermal resistance of the couple
E_0 = open-circuit voltage of the couple
r = coefficient of the electric current in the expression for the terminal voltage
s = coefficient of the square of the current in the expression for the terminal voltage
η = conversion efficiency of the generator
η_0 = optimum conversion efficiency
I_0 = optimum load current
Z = figure of merit of the couple
$\bar{T}$ = average of the heat source and heat sink temperatures
M = ratio of optimum load resistance to generator resistance for the zero thermal resistance case

REPRESENTATIVE PARAMETER VALUES

$l = 1$ cm (centimeter)
$A_1 = A_2 = 1$ cm^2
$k_1 = k_2 = 1.5 \times 10^{-2}$ watts/cm-K
$\rho_1 = \rho_2 = 10^{-3}$ ohm-cm
$-\alpha_1 = \alpha_2 = 200$ microvolts/K
$\alpha_l = \alpha_0 = 400$ microvolts/K
$T_h = 600$ K
$T_c = 300$ K
$Z = 2.66 \times 10^{-3}$ (K)$^{-1}$

References

1. SEMICONDUCTOR THERMOELEMENTS AND THERMOELECTRIC COOLING (book), A. F. Ioffe. Infosearch Ltd., London, England, 1957.

2. THE EFFICIENCY OF THERMOELECTRIC GENERATORS, J. M. Borrego, H. A. Lyden, J. Blair. *Technical Note 58-200*, Wright Air Development Center, Dayton, Ohio, Project 6058, Nov. 1958.

3. MULTIPLE STAGE THERMOELECTRIC GENERATION OF POWER, T. C. Harman. *Journal of Applied Physics*, New York, N. Y., vol. 29, 1958, pp. 1471–73.

The Effects of Nuclear Radiation on the Electrical Strength of Air

G. I. DUNCAN
MEMBER AIEE

J. C. FRASER
MEMBER AIEE

B. VALACHOVIC
NONMEMBER AIEE

WITH THE EMPHASIS today upon extended environmental conditions for electronic equipment used in military aircraft and guided missiles, there is a pressing need for the simultaneous testing of such equipment under two or more of these extreme environments. This paper describes a program completed by the General Electric Company covering the testing of spark gaps in the Brookhaven National Laboratory's graphic reactor. It consisted of a 2-week in-pile exposure designed to investigate the voltage breakdown strength of air at various pressures in the presence of intense nuclear radiation of the following approximate levels:

Fast flux: 1×10^{11} fast neutrons per cm^2-sec (centimeter^2-seconds)
Thermal flux: 2×10^{12} thermal neutrons per cm^2-sec
Gamma flux: 1×10^{13} gamma photons per cm-sec^2

Present environmental goals as established by the U.S. Air Force for electronic power transformers are 100,000-foot altitude, 500 C (degrees centigrade) ambient temperature, and intense nuclear radiation, approximating the levels previously mentioned. Since terminal spacings have formerly been designed for air electric strengths to approximately 50,000 feet and temperatures up to 125 C, additional information was required on breakdown voltages under these new conditions.

Previous work on determining the voltage spacing characteristics as a function of air density up to altitudes of 150,000 feet[1] has not included high temperature or radiation. A search of the literature revealed little or no information on the subject. Consequently, an experiment was designed and conducted to obtain this information.

Test Variables and Levels

The electrical breakdown strength of air is dependent upon air density, electrode spacing, electrode material, electrode configuration, frequency of voltage, and rate of increase of voltage.

To gain the maximum amount of information, it was decided to obtain data for three levels of gap spacing and three levels of air pressure, with and without radia-

---

Paper 59-134, recommended by the AIEE Electrical Insulation Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 1–6, 1959. Manuscript submitted November 25, 1958; made available for printing October 7, 1959.

G. I. DUNCAN is with the General Electric Company, Fort Wayne, Ind., and J. C. FRASER and B. VALACHOVIC are with the General Electric Company, Schenectady, N. Y.

The work described in this paper was part of Air Force Contract No. *AF33(616)-5579* for developing temperature- and radiation-tolerant electronic power transformers.

We are deeply indebted to the staff members of the Brookhaven National Laboratory Reactor Department, and in particular to J. J. Floyd, G. C. Kinne, R. W. Powell, and F. P. Reeve, for their co-operation and assistance in the performance of our radiation testing program. We also want to acknowledge the work done by T. W. Moller of the General Electric General Engineering Laboratory in designing and building the dynamic pressure system used for these tests.

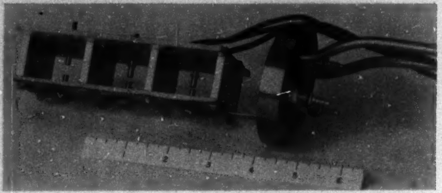


Fig. 1. Electrode assembly

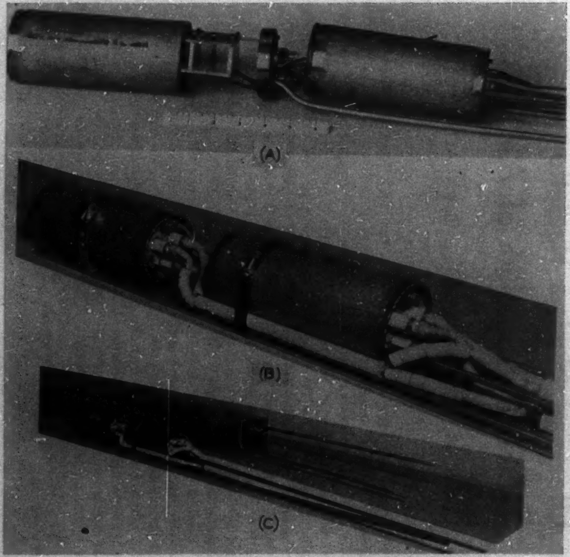


Fig. 2. Spark gap items. A—Assembly. B—Test specimen and tray. C—Test specimens and container

Fig. 3 (right). Circuit for spark gap breakdown tests

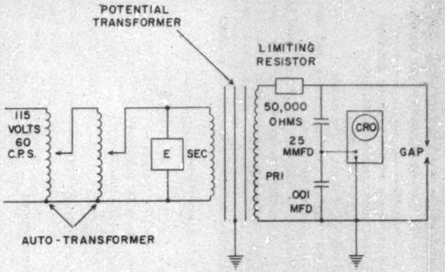
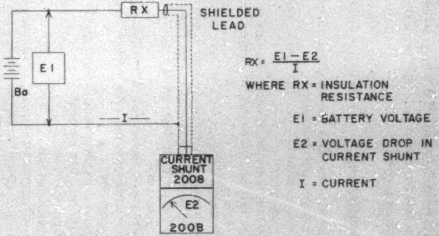

POTENTIAL TRANSFORMER
LIMITING RESISTOR
115 VOLTS 60 C.P.S.
E
SEC
50,000 OHMS
25 MMFD
CRO
GAP
PRI
.001 MFD
AUTO-TRANSFORMER

Fig. 4 (right). Circuit for gap insulation resistance tests


RX
SHIELDED LEAD
E1
Ba
I
CURRENT SHUNT 2008
E2
200B

$RX = \frac{E1 - E2}{I}$

WHERE RX = INSULATION RESISTANCE

E1 = BATTERY VOLTAGE

E2 = VOLTAGE DROP IN CURRENT SHUNT

I = CURRENT

Fig. 5 (right). Dynamic pressure system

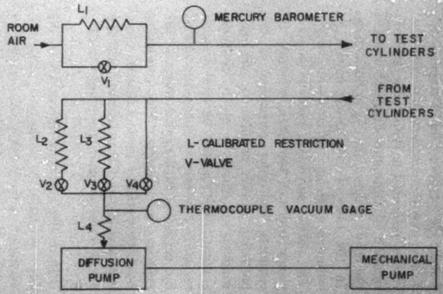

ROOM AIR
L1
V1
MERCURY BAROMETER
TO TEST CYLINDERS
FROM TEST CYLINDERS
L2
L3
V2
V3
V4
L-CALIBRATED RESTRICTION
V-VALVE
THERMOCOUPLE VACUUM GAGE
L4
DIFFUSION PUMP
MECHANICAL PUMP

Table I. Spark Gap Breakdown Tests, No Radiation

Cylinder 2

Container Ambient Temperature, C	Cylinder Pressure, Mm Hg	Breakdown Voltage
1/16-Inch Gap		
26	756.2	3,780 3,900 3,870 3,880 3,889
	75.0	960 1,020 980 1,010 970
26	7.0	366 372 366 384 378
77	763.3	3,810 3,750 3,570 3,690
76	71.0	852 876 895 864
75	7.4	363 364 366 358
128	766.1	3,450 3,420 3,470 3,450
	70.0	805 823 810 823
129	7.0	348 324 318 315
168	763.8	3,240 3,210 3,210 3,180
	76.0	834 834 816 822
169	7.9	339 321 318 321

Container Ambient Temperature, C	Cylinder Pressure, Mm Hg	Breakdown Voltage
1/8-Inch Gap		
26	756.2	4,680 4,800 4,870 4,920 4,800
	75.0	1,470 1,480 1,440 1,470 1,475
	7.0	720 540 510 498 510
77	763.3	4,500 4,620 4,620 4,710
76	71.0	1,260 1,315 1,270 1,260
	7.4	456 457 468 474
128	766.1	4,200 4,290 4,260 4,410
	71.0	1,200 1,260 1,190 1,200
	7.6	441 438 435 428
168	763.8	4,050 4,050 4,050 4,050
	76.0	1,200 1,210 1,190 1,200
169	7.0	417 432 417 420

Container Ambient Temperature, C	Cylinder Pressure, Mm Hg	Breakdown Voltage
1/4-Inch Gap		
26	656.2	6,240 6,350 6,300 6,490 6,600
	75.0	2,000 1,830 1,800 1,740 1,740
	7.0	696 624 624 618 624
77	763.2	5,640 5,820 5,940 5,820
76	71.0	1,560 1,560 1,520 1,510
	7.4	654 600 573 589
128	766.1	5,040 4,860 4,970 4,970
	71.0	1,380 1,440 1,440 1,385
	7.6	546 549 555 546
168	763.8	4,680 4,680 4,740 4,860
	76.0	1,360 1,350 1,350 1,370
	7.9	534 510 528

Fig. 6. Reactor mock-up

Table II. D-c Insulation Resistance, No Radiation

Cylinder 2

Container Ambient Temperature, C	Cylinder Pressure, Mm Hg	Insulation Resistance, Ohms		
		1/16-Inch Gap	1/8-Inch Gap	1/4-Inch Gap
23	754.0	1×10^{11}	1×10^{11}	1×10^{11}
	76.0	1×10^{11}	1×10^{11}	2×10^{11}
	7.4	9×10^{10}	1×10^{11}	1×10^{11}
77	7.4	4×10^{11}	7×10^{11}	6×10^{11}
125	7.6	1×10^{11}	9×10^{10}	1×10^{11}
169	7.6	4×10^{4}	2×10^{10}	2×10^{10}

Table III. Spark Gap Breakdown Tests, During Radiation

Cylinder 2

Container Ambient Temperature, C	Cylinder Pressure, Mm Hg	Breakdown Voltage	Container Ambient Temperature, C	Cylinder Pressure, Mm Hg	Breakdown Voltage	Container Ambient Temperature, C	Cylinder Pressure, Mm Hg	Breakdown Voltage
	1/16-Inch Gap			1/8-Inch Gap			1/4-Inch Gap	
100	750.0	2,880 2,880 2,928 2,970	100	750.0	4,320 4,350 4,308 4,140	100	750.0	7,050 6,990 6,990 6,900
	82.0	816 822 822 828		80.5	1,092 1,098 1,095 1,092		80.0	1,254 1,242 1,248 1,242
	23.0	432 426 429 435		24.0	588 588 591 597		22.0	696 693 696 696
122	745.5	2,670 2,682 2,682 2,688	122	745.5	4,038 4,020 3,990 4,038	122	745.5	6,720 6,600 6,570 6,510
126	79.0	774 765 774 774	126	81.0	1,062 1,062 1,056 1,050	127	83.0	1,188 1,182 1,185 1,185
130	25.5	427 429 432 435	130	26.0	594 594 594 594	130	27.0	738 741 741 744
147	753.5	2,430 2,442 2,400 2,412	147	753.5	2,580 2,580 2,598 2,580	147	751.8	3,300 3,270 3,270 3,270
146	74.5	702 702 708 708	146	78.0	870 870 876 876	146	79.0	900 948 936 948
147	27.0	420 426 423 414	147	26.0	528 528 528 528	147	25.0	510 492 486 492
153	762.5	2,712 2,694 2,700 2,700	153	762.5	2,970 2,976 3,060 3,090	153	760.5	2,580 2,610 2,610 2,616
	74.0	690 690 702 796		75.0	894 888 882 882		76.5	966 972 978 978
	20.5	360 360 366 366		22.0	482 482 482 494		24.0	480 480 480 480
	7.5	258 264 264 264		7.5	318 324 327 330		7.5	348 336 360 354

tion. The gap spacings selected were 1/16 inch, 1/8 inch, and 1/4 inch. Rod electrodes were selected for the gaps as it was felt that they would more closely approach the nonuniform fields normally encountered with terminals than would other types of gaps.

Air pressure was selected as a variable instead of air density because in free space the air pressure is not affected by changes in temperature, whereas temperature does affect air density. Air pressure can be directly related to altitude independent of temperature. The air pressures originally selected were 760, 76, and 7.6 mm (millimeters) of mercury which correspond roughly to sea level, 50,000-foot altitude, and 100,000-foot altitude respectively.

Voltage breakdown and d-c conductivity measurements were made on the spark gaps in a simulated reactor mock-up for the no-radiation condition, and in the Brookhaven National Laboratory for radiation conditions.

Because of random variations in breakdown voltage, a spark gap can be expected to give only approximate values. To obtain reasonably correct values of breakdown voltage, the average of a number of measurements was taken.

Test Specimen

Two independent, identical test specimens were built and used to insure test reliability; see Figs. 1 and 2. Each specimen consisted of an electrode holder having three gaps in a hermetically sealed stainless steel cylinder. The gap electrodes were made from 1/8-inch-diameter nickel spark plug electrodes, the faces of

Fig. 7. Test location at Brookhaven National Laboratory

Table IV. D-c Insulation Resistance, During Radiation

Cylinder 2

Container Ambient Temperature, C	Cylinder Pressure, Mm Hg	Insulation Resistance, Ohms
	1/16-Inch Gap	
100	750.0	5×10^7
	67.0	6×10^7
	21.0	7×10^7
122	745.5	4×10^7
126	84.5	5×10^7
130	22.0	
147	753.5	3×10^7
146	56.5	4×10^7
147	25.0	
153	762.5	2×10^7
	79.0	3×10^7
	22.0	
	1/8-Inch Gap	
100	750.0	5×10^7
	68.0	6×10^7
	21.0	7×10^7
122	745.5	4×10^7
126	84.5	5×10^7
130	22.0	
147	753.5	2×10^7
	77.5	4×10^7
	26.0	
153	762.5	2×10^7
	79.0	3×10^7
	22.0	
	1/4-Inch Gap	
100	750.0	5×10^7
	68.0	
	21.0	7×10^7
122	745.5	4×10^7
126	78.5	
130	22.0	5×10^7
147	753.5	2×10^7
146	77.5	3×10^7
147	26.0	4×10^7
153	762.5	2×10^7
	30.0	3×10^7
	22.0	

which were machined flat to eliminate rounded edges. They were held in place by sheets of 1/8-inch-thick aluminum phosphate bonded muscovite mica separated by 1/4-inch-thick aluminum spacers. All electrodes on one side of the assembly, and the aluminum spacers, were connected to a stainless steel ground strap, which was resistance-welded to the front end of the test cylinder. The mica sheet on the high-voltage side of the assembly was threaded for ease in adjusting the gaps; gage blocks were used to set the gaps.

The test cylinder for each assembly consisted of a 3-inch outside diameter, a 0.065-inch-thick 7-inch-long stainless steel cylinder, with a stainless steel cap welded into each end. Two wall tubes, 1/4-inch in outside diameter, 0.035-inch thick, for the dynamic pressure system were arc-welded into the front end cap. Three alumina ceramic hermetic terminals were brazed into the front cap with the use of a 5% silver, 95% cadmium brazing alloy. The high-voltage leads for the gaps were brought out of the cylinder through these terminals. A small stainless steel screw was projection-welded to the front cap for use as a ground terminal. Both caps were arc-welded into the ends of the stainless steel cylinder.

Electric connections between the high-voltage electrodes and the hermetic terminals were made with 1/32-inch silicone rubber-insulated American Wire Gage no. *18* solid copper conductor, using mechanical connectors at the electrodes, and silver brazing to the terminals. The high-voltage leads from the terminals to the outside wall of the reactor consisted of a short length of the copper wire described above, brazed to 1/16-inch aluminum conductors insulated with a 3/64-inch wall of silicone rubber.

Instrumentation

Because of the inaccessibility of the spark gap cylinders in the reactor, remote instrumentation was required. Fig. 3 shows the circuit used for the breakdown tests. A conventional instrument-type potential transformer was used in reverse to supply the high voltage. The normal transformer secondary was used as the input in this application and a General Electric Company type *P3* voltmeter was used to measure the input voltage. In the output side of this potential transformer, which is normally the primary side, there was a limiting resistor to prevent excessive load on the transformer when the spark gap should break down.

A cathode-ray oscilloscope was used to

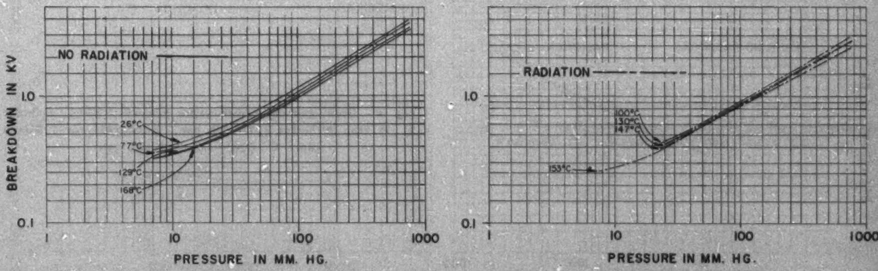


Fig. 8. Spark gap breakdown versus pressure for 1/16-inch gap. A (left)—With no radiation. B (right)—During radiation

indicate breakdown through the change in spark gap voltage waveshape. The voltage control on the input side of the transformer consisted of two variable voltage autotransformers connected to give coarse and fine control.

The d-c conductivity measurements were made using a Keithley d-c vacuum-tube voltmeter of the electrometer type as the basic measuring element. This is a battery-operated unit which measures direct voltage directly; it measures current in conjunction with a decade shunt. These measurements were made at approximately 90 volts direct current. A schematic diagram of the circuit used for insulation resistance measurements is shown in Fig. 4.

Temperature measurements were made using a chromel-alumel thermocouple located in the test container midway between the two test cylinders. Because of the gamma heating effect upon cylinder materials, particularly the stainless steel, test cylinder temperatures increased at a gradual rate throughout the test up to 153 C. Temperature measurements made in the *E-52* hole previous to the start of this test showed a hole temperature of approximately 70 C.

Dynamic Pressure System

To establish and to measure the pressure necessary to simulate the various altitudes desired in the spark gap breakdown tests, a dynamic pressure system was selected. The reasons for this selection were twofold. First, at the low pressure required to simulate a 100,000-foot altitude (approximately 7.6 mm of mercury), water vapor and other volatiles within the test cylinder would cause a drift in pressure if the system were a static one. Second, under both the 50,000-foot and 100,000-foot altitude conditions, the absorbed gases driven off the electrode surfaces at each firing would necessitate constant readjustment of the pressure within a static system. With the dynamic system, equilibrium conditions at the desired pressures were easily established and recovery after transient gas surges was automatic. A continuous flow of room air through the test cylinder also served to flush out any contaminants within the cylinder which might tend to modify the test results.

Fig. 5 shows a schematic diagram of the system used in the spark gap tests. Calibrated restrictions on the inlet and outlet sides of the cylinder were used for each pressure required below atmospheric. For atmospheric measurements in the cylinders, the outlet valves (*V2*, *V3*, and *V4*) were closed, and an inlet by-pass valve (*V1*) opened, so that the vacuum system could continue "pump-down" while atmospheric measurements were being taken. The two test cylinders were connected in parallel during pump-down. The arrangement of pressure valves permitted selection of either cylinder for measurements, as desired.

Simulated Reactor Mock-up Tests

To obtain breakdown voltage and d-c conductivity data on the spark gaps without radiation, a full-scale mock-up of a reactor hole was constructed. Fig. 6 shows a view of this mock-up. It consisted of an aluminum inner duct 1/4 inch thick, 4 by 4 inches inside dimensions, and 18 feet long, to simulate the 4-inch-square test hole in the Brookhaven reactor. By means of strip heaters surrounding the duct, it could be heated to temperatures as high as 200 C, so that reactor hole ambient temperatures could be duplicated.

Vol.
Pt. 1
196

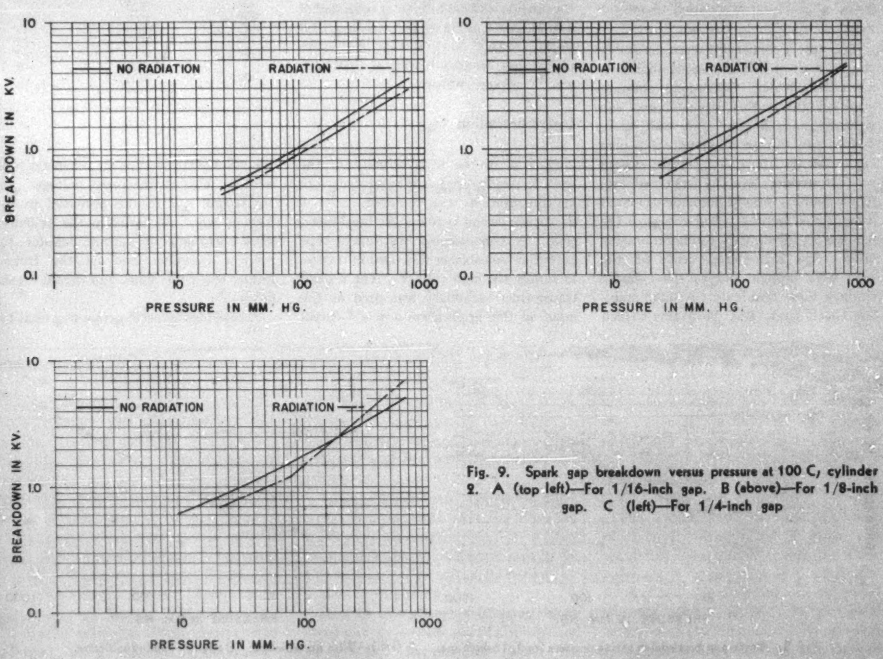


Fig. 9. Spark gap breakdown versus pressure at 100 C, cylinder 2. A (top left)—For 1/16-inch gap. B (above)—For 1/8-inch gap. C (left)—For 1/4-inch gap

Sixty-cycle-per-second breakdown voltage and d-c conductivity tests were performed with the spark gap specimens inserted into the reactor mock-up. Measurements were made at approximately 26, 75, 129, and 169 C, and at approximately 760, 76, and 7.6 mm of mercury.

The data for one of the test cylinders are shown in Tables I and II.

Radiation Tests

The radiation testing was carried on at Brookhaven National Laboratory in the *E-52* hole. Fig. 7 shows the test specimens being inserted into the reactor.

For most of the test, the reactor was operated at 13 megawatts. The test specimens were in the reactor for a total elapsed time of approximately 330 hours. During this test, measurements of breakdown voltage, d-c conductivity, and temperature were made periodically at approximately 760, 76, and 7.6 mm of mercury.

The data for one of the test cylinders are shown in Tables III and IV.

Test Results

To show the effects of nuclear radiation on the electric strength of air, a family of temperature curves was plotted as a function of breakdown voltage and pressure for each gap and condition of radiation. The curves for the 1/16-inch gap in one of the test cylinders are shown in Figs. 8(A) and (B).

From these curves, breakdown voltage versus pressure data were interpolated for specific temperatures. Fig. 9 shows these data plotted for 100 C.

Fig. 10 shows the d-c insulation resistance as a function of pressure for the gaps in one of the test cylinders for both the no-radiation and radiation conditions. The no-radiation data are shown only at 23 C.

Conclusions

The data obtained in the tests and the curves plotted from these data indicate that nuclear radiation at the high intensities investigated had a noticeable effect on breakdown voltage and d-c conductivity of a gap. The data also demonstrated, as expected, that the breakdown voltage of a gap decreases with decreasing pressure (temperature constant) and increasing temperature (pressure constant).

Test results show that under conditions of intense nuclear radiation, spark gap breakdown is lowered about 20%. This is probably due to the effect of photoelectrons during radiation. This effect makes available a large number of electrons in the gap at all times to start ionization, resulting in increased conductivity, and breakdown at lower voltage values.

Earlier investigators of spark gap breakdown found that accuracy could be increased by irradiating the gaps with ultraviolet light, with a resulting reduction in the average value of breakdown voltage by several per cent. It is to be expected that nuclear radiation would have an appreciably greater effect. Since radiation is a form of energy, the probability of photoionization is proportional to the radiation intensity.

There is no logical explanation for the

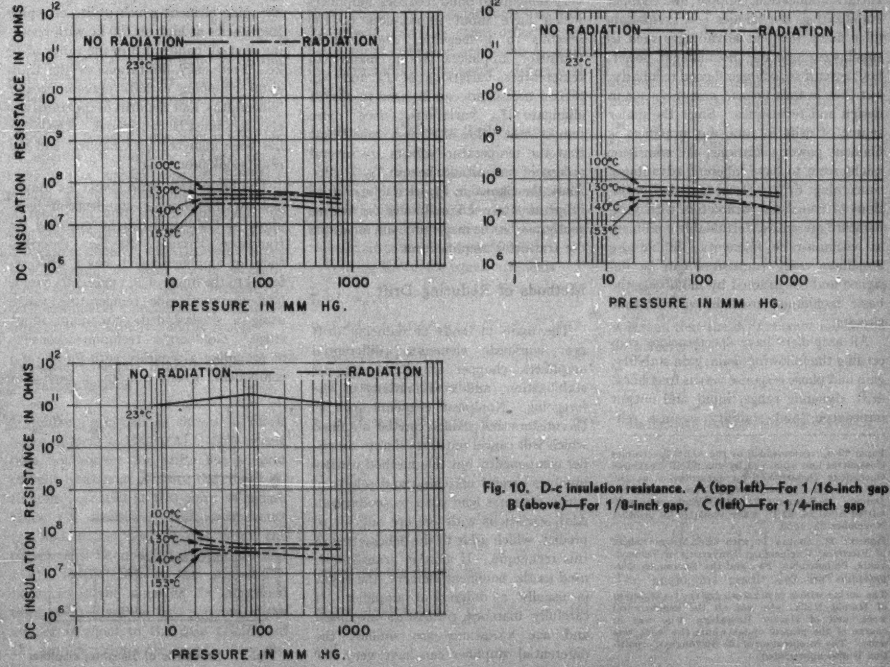


Fig. 10. D-c insulation resistance. A (top left)—For 1/16-inch gap B (above)—For 1/8-inch gap. C (left)—For 1/4-inch gap

crossover of the curves in Fig. 9(C). A close examination of the data under the heading "1/4-Inch Gap" in Table III suggests that the data for the 1/4-inch gap during radiation may be suspect.

Radiation accounted for at least a 3-decade lowering in gap d-c insulation resistance. Changes in d-c insulation resistance, under radiation conditions, as a result of varying temperatures, were less than half a decade.

Knowledge in the field of radiation effects is still sparse. The amount of general information on components under radiation is particularly limited. It is hoped that the experimental data obtained in these tests will serve to shed some light upon the theories advanced, and will be helpful in evaluating and conducting subsequent tests on all types of electronic components. These tests will serve as the basis for a designer to establish safe voltage breakdown distances between terminals, and between terminals and ground, for transformer applications under extreme altitude and radiation conditions.

Reference

1. Terminal Spacings for High-Altitude and Ultrahigh-Temperature Electronic Transformer Applications, G. I. Duncan, W. A. Rectanus. *AIEE Transactions*, pt. II (*Applications and Industry*), vol. 78, Mar. 1959, pp. 16–19.

Stable Transistor Wide-Band D-C Amplifiers

ROBERT H. OKADA
NONMEMBER AIEE

STABLE D-C amplification is required in many engineering applications. In the past, vacuum-tube circuits have been developed which meet most low drift requirements by utilizing differential amplification, chopper, and chopper stabilization techniques. Transistors offer many inherent advantages such as small size, long life, no filament power, low power dissipation and good reliability, but they in turn present new problems in design and techniques. Since the major source of drift in tube d-c amplifiers is filament power variations, the transistor might seem to have inherent advantages concerning drift. However, the variations of transistor parameters with temperature presents as formidable a problem as vacuum-tube filaments. Stable d-c amplifiers using transistors can be designed and constructed by modifying the basic techniques used in vacuum-tube circuits.

All amplifiers have specifications concerning the following: gain, gain stability, gain and phase response versus frequency, drift, dynamic range, input and output impedance, and stability against self-oscillation if feedback is used. Transistors differ from tubes notably in the impedance levels and d-c drift. Although most transistor circuit configurations present a low input impedance, it is not very difficult to build feedback amplifiers which have input impedances of 100 kilohms to 1 megohm. D-c drift in transistor amplifiers is due mainly to temperature variations of I_{co} and V_{be}. Silicon transistors can be used to almost eliminate I_{co} variations, since room temperature value of I_{co} is so small that that the temperature effects are several orders of magnitude below V_{be} effects. Thus, the change in V_{be} with temperature (approximately 2.5 millivolts per degree centigrade) is the major problem for stable d-c transistor amplification.

Methods of Reducing Drift

The main methods of reducing drift are: nonlinear elements, differential amplifiers, chopper amplifiers, chopper stabilization, and combinations of the foregoing. Nonlinear elements such as thermistors and diodes can be obtained which will cancel out drift due to amplifier components, but this method usually involves careful matching and selection, which does not lend itself to production. Also, variations with age are difficult to predict, which adds to the delicateness of this technique. If another transistor is used as the nonlinear element, the result is usually a differential amplifier. If carefully matched transistors are used, and age variations are similar, the differential amplifier can have very low drift. Slaughter presented a differential amplifier which had excellent drift characteristics (due to well-matched transistors).[1] Fig. 1 is a schematic of this amplifier.

Chopper amplifiers, which modulate the input signal at a fixed frequency, amplify the resulting carrier and side bands, and then demodulate, are essentially drift-free, provided the chopper itself is a stable element. Mechanical choppers can be used which introduce negligible drift but are limited in frequency to about 1 kc. The over-all bandwidth of such an amplifier is a few hundred cycles. Transistors can be used as the chopping element which utilize chopping frequencies as high as 100 kc with resulting over-all bandwidth in the range of 10 kc. Frequency limitations of the transistor chopping circuits prevent greater bandwidths, and the drift of the transistor choppers must be considered. The latter has been analyzed by Chaplin,[2] and the results look promising.

To obtain wide-band stable d-c amplification the chopper stabilization technique of Goldberg[3] can be modified for transistor circuits. Blecher describes such an amplifier which has a drift referred to the input of ±5 millivolts over a 50-degree-centigrade temperature range using a single-ended germanium input stage.[4] Goldberg's technique consists of providing a separate path for the d-c component, which contains a high-gain

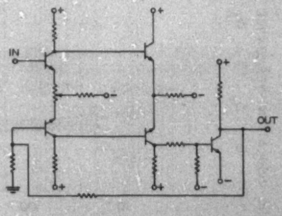


Fig. 1. Schematic of Slaughter amplifier

Paper 60-6, recommended by the AIEE Electronics Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., January 31–February 5, 1960. Manuscript submitted April 24, 1959; made available for printing November 23, 1959.

Robert H. Okada is with the Moore School of Electrical Engineering, University of Pennsylvania, Philadelphia, Pa., and the Burroughs Corporation, Paoli, Pa.

The author wishes to acknowledge the contribution of Harold Noffz, who did all the experimental work, and of Harvey Rosenberg, who was in charge of the project under which the work was done. The co-operation of the Burroughs Corporation is also appreciated.

Vol.
Pt. 1
196

low-drift chopper amplifier preceding the main amplifier.

By a combination of silicon transistor differential amplifier and chopper stabilization, a transistor amplifier can be built with the use of transistors which have normal production spreads of V_{be} temperature variations, and resulting in a drift referred to input of less than one half of a millivolt. By selecting matched V_{be} temperature characteristics the drift can be reduced much farther.

Transistor Operational Amplifiers

The standard voltage operational amplifier is shown in Fig. 2(A) with A_v a voltage gain (negative). The gain of the amplifier is Z_f/Z_i if $|A_v|>>1$ and $Z_1>>(Z_f/A_v)$. For a transistor amplifier the last requirement is usually not true. However, consider the current amplifier in Fig. 2(B). From the Appendix is obtained

$$\frac{I_L}{I_i} \doteq -\frac{Z_f}{Z_L} \text{ if } Z_f \gg \frac{Z_1}{A_i} \text{ and } Z_L \gg \frac{Z_f+Z_1}{1+A_i}$$

and:

$$\frac{E_L}{E_i} = \frac{-Z_f}{Z_i} \text{ if } A_i\beta_i \gg 1,\ A_i\beta_i \gg \frac{2Z_1}{Z_i}, \text{ and}$$

$$\beta_i \equiv \frac{I_f}{I_L} \doteq \frac{Z_L}{Z_f}$$

Thus, for the voltage gain to be the ratio of two impedances, the input impedance at the A_i amplifier terminals should ideally be zero, and a low input impedance transistor amplifier is ideal for an operational amplifier. These equations are a special case of Blecher's results which utilize the short-circuit current gain and source impedance in place of a current amplifier.[4] However, if it is remembered that the value of A_i is a function Z_L (and Z_f in general), these equations are useful because of their simplicity, especially since the above conditions are usually valid in practice.

Possible Chopper Stabilization Configurations

Goldberg's circuit is shown in block diagram form in Fig. 3(A). This circuit does not correct for drifts due to grid current leakage at A since the circuit interprets potential at this point as signal. Since the transistor drift components (I_{co} and V_{be}) appear at this point, the circuit must be modified as shown in Fig. 3(B). If the ratio of the primed impedances is the same as the unprimed ones, point B will be zero potential if the output is equal to the gain (E_0/E_i) of the amplifier times the input. Any discrepancy is due to drift components and the chopper amplifier will attempt to correct for the drift. Incidentally, aside from drift signals appearing at point A, point B is at the same potential as point A, and therefore the primed impedances can be removed and the two points A and B tied together. This reduces Fig. 3(B) to Fig. 3(A) and is an excellent means of explaining to the uninitiated the (at first) puzzling configuration of Fig. 3(A).

The chopper-stabilizing effects are limited (as will be shown later), which makes a differential input stage desirable in order to reduce the amount of stabilization required. In a typical example the V_{be} drift over a range of 70 F (degrees Fahrenheit) to 150 F will be 100 millivolts, which is to be compared with a ΔV_{be} in a differential input stage of 17 millivolts. The 17-millivolt figure is the maximum ΔV_{be} variation of 32 silicon transistors, type *2N479*, over the same temperature range. Thus if these figures are typical of production results, the differential input amplifier has an inherent drift reduction of about 5 to 1 without selection. The differential stage can also supply the high common mode rejection necessary if a single-ended output is required. If silicon transistors are used, the I_{co} effects are negligible compared with V_{be} effects, but the same order of inherent drift reduction applies to I_{co} drift in germanium transistor differential stages.

Analysis of Chopper Stabilization

Fig. 3(C) represents the same circuit as 3(B), in which A_i and A_i' represent actual current gains of amplifiers *G*1 and *G*2, and I_k is the current required to reduce the output to zero because of drift.

Writing Kirchoff's equations yields

$$I_0+I_0'=I_{in} \tag{1}$$

$$-(I_1+I_0+I_f+I_k)A_i=I_f+I_f'+I_L \tag{2}$$

$$-(I_0'+I_f')A_i'=I_1 \tag{3}$$

and if (see the Appendix)

$$Z_L \gg \frac{Z_{in}}{A_i},\ Z_f \gg \frac{Z_{in}}{A_i},\ Z_L \gg \frac{Z_{in}'}{A_i'}$$

and

$$Z_f' \gg \frac{Z_{in}'}{A_i'}$$

$$Z_LI_L=Z_fI_f=Z_f'I_f'$$

and if (see the Appendix)

$$A_i \gg 1,\ A_i' \gg 1,\ Z_i \gg \frac{Z_{in}}{A_i\beta_i} \tag{4}$$

and

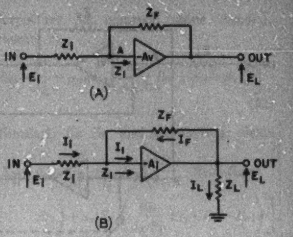


Fig. 2. Operational amplifiers. A—Voltage mode. B—Current mode

$$Z_f' \gg \frac{Z_{in}'}{A_i'\beta_i'}$$

then

$$Z_iI_0=Z_i'I_0' \tag{5}$$

Solving equations 1 and 5 simultaneously yields the load current in terms of the input current

$$I_L=-\frac{Z_f'}{Z_L}\frac{Z_i}{Z_i+Z_i'}\left[I_{in}+\frac{Z_i+Z_i'}{Z_i}\frac{I_k}{A_i'}\right] \tag{6}$$

if, as is usually the case,

$$A_i \gg 1,\ A_i' \gg \frac{1}{A_i},\ \frac{A_i'}{Z_f'} \gg \frac{1}{Z_f}+\frac{1}{AZ_L'},\ \frac{A_i'Z_i}{Z_i'} \gg 1$$

If equation 6 is compared with the analogous result using no chopper stabilization (see the Appendix), i.e.,

$$I_L=-\frac{Z_f}{Z_L}(I_{in}+I_k)$$

if

$$Z_f \gg \frac{Z_{in}}{A_i}$$

and

$$Z_L \gg \frac{Z_f+Z_1}{1+A_i} \tag{7}$$

it is seen that the drift current is reduced by a factor of

$$\frac{Z_i+Z_i'}{Z_i}\frac{1}{A_i'}$$

In terms of voltage, since in Fig. 3(C)

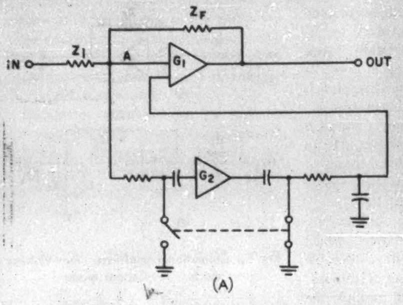
$$E_L=I_LR_L,\ E_{in} \doteq I_{in}\frac{Z_iZ_i'}{Z_i+Z_i'}$$

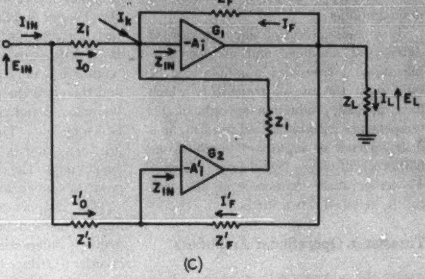
the output voltage is

$$E_L=-\frac{Z_f'}{Z_i'}\left[E_{in}-\frac{Z_i'I_k}{A_i'}\right] \tag{8}$$

The analogous result for no chopper stabilization is

$$E_L=\frac{-Z_f}{Z_i}[E_{in}-Z_iI_k] \tag{9}$$


ZF
IN
Z1
A
G1
OUT
G2
(A)


IIN
EIN
Zi
Io
Ik
ZF
IF
-Ai
G1
ZIN
ZL
IL
EL
Zi
G2
-A'i
ZIN
I'o
I'F
Z'i
Z'F
(C)

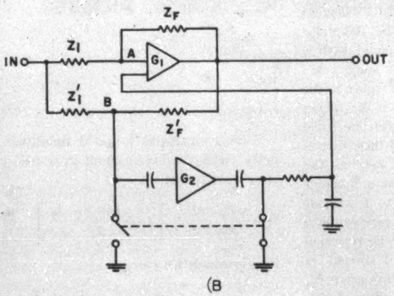

ZF
IN
Z1
A
G1
OUT
Z'1
B
Z'F
G2
(B

Fig. 3. A—Schematic of Goldberg's chopper stabilization scheme. B—Modification for transistor circuits. C—Schematic of B used for determining drift

and the ratio of the drift terms is

$$\frac{\dfrac{Z_i' I_k}{A'}}{Z_i I_k} = \frac{Z_i'}{Z_i A_i}$$

or voltage drift reduction is

$$\frac{Z_i'}{Z_i}\,\frac{1}{A_i'} \qquad (10)$$

The choice of the ratio Z_i'/Z_i is determined by the required input impedance which is equal to the parallel combination of Z_i and Z_i'. The chopper current gain A_i' can be very high and in mechanical choppers is limited only by the electrostatic and electromagnetic pickup at the vibrating contracts, and the minimum value of Z_1 in Fig. 3(C).

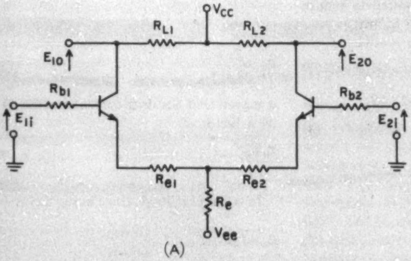
Analysis of Differential Amplifiers

A differential amplifier is shown in Fig. 4(A). To study the effects of I_{co}, V_{be}, and dissimilar transistors, the equivalent circuit of Fig. 4(B) will be used. (This circuit treats the base-to-emitter voltage as an externally available voltage which can be measured experimentally. Both d-c and varying signals are included. The basic assumptions for this equivalent circuit are: 1. The constant α_n. 2. The collector impedance is very high compared with load impedance. 3. Any internal base resistance is included in R_{b1} and R_{b2}.) The branch equations are

$$E_1 - V_{be1} - V_{ee} = R_{b1}I_{b1} + (R_{e1}+R_e)I_{e1} + R_eI_{e2} \qquad (11)$$

$$E_2 - V_{be2} - V_{ee} = R_{b2}I_{b2} + R_eI_{e1} + (R_{e2}+R_e)I_{e2} \qquad (12)$$

$$I_{e1} = I_{b1} + \alpha_{n1}I_{e1} + I_{co1} \qquad (13)$$


Vcc
RL1
RL2
E1o
E2o
Rb1
Rb2
E1i
E2i
Re1
Re2
Re
Vee
(A)

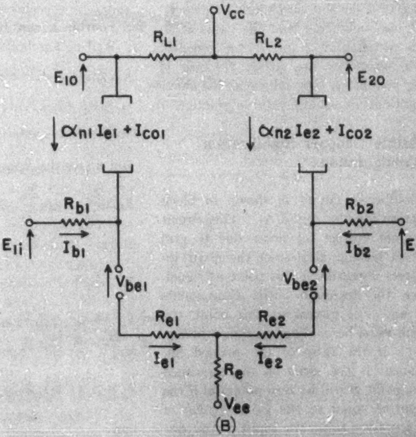

Vcc
RL1
RL2
E1o
E2o
αn1 Ie1 + Ico1
αn2 Ie2 + Ico2
Rb1
Rb2
E1i
Ib1
Ib2
E2i
Vbe1
Vbe2
Re1
Re2
Ie1
Ie2
Re
Vee
(B)

Fig. 4. A—Transistor differential amplifier stage. B—An equivalent circuit for A

$$I_{e2}=I_{b2}+\alpha_{n2}I_{e2}+I_{co2} \qquad (14)$$

Using the relation, $\beta_n=\alpha_n/1-\alpha_n$, and solving equation 13 for I_{b1} yields

$$I_{b1}=I_{e1}\left(\frac{1}{1+\beta_{n1}}\right)-I_{co1} \qquad (15)$$

Similarly,

$$1_{b2}=I_{e2}\left(\frac{1}{1+\beta_{n2}}\right)-I_{co2} \qquad (16)$$

Equations 11, 12, 15, and 16 can be solved for I_{e2}, with the use of the knowledge that

$$E_{2o}=V_{cc}-(\alpha_{n2}I_{e2}+I_{co2})R_{L2} \qquad (17)$$

The result is

$$E_{2o}=V_{cc}-\alpha_{n2}R_{L2}\times\left[\frac{(E_{2i}-E_{1i})+(V_{be1}-V_{be2})+R_{b2}I_{co2}-R_{b1}I_{co1}-\left(\frac{R_{b1}}{\beta_{n1}}+R_{e1}\right)\frac{V_{ee}}{R_e}}{R_{e1}+R_{e2}+\frac{R_{b2}}{\beta_{n2}}+\frac{R_{b1}}{\beta_{n1}}+\frac{R_{b1}R_{b2}}{\beta_{n1}\beta_{n2}R_e}}\right]-I_{co2}R_{L2} \qquad (18)$$

provided $R_{b1}/\beta_{n1}R_e\ll 1$. A similar result is obtained for E_{1o} and, taking the difference,

$$\begin{aligned}E_{2o}-E_{1o}={}&\frac{(E_{1i}-E_{2i})(\alpha_{n1}R_{L1}+\alpha_{n2}R_{L2})}{\Delta}+\\&\frac{(V_{be2}-V_{be1})(\alpha_{n1}R_{L1}+\alpha_{n2}R_{L2})}{\Delta}+\\&\frac{(R_{b1}I_{co1}-R_{b2}I_{co2})}{\Delta}\times\\&\frac{(\alpha_{n1}R_{L1}+\alpha_{n2}R_{L2})}{1}+\\&\left(\frac{R_{b1}}{\beta_{n1}}+R_{e1}-\frac{R_{b2}}{\beta_{n2}}-R_{e2}\right)\times\\&\frac{\frac{V_{ee}}{R_e}(\alpha_{n1}R_{L1}+\alpha_{n2}R_{L2})}{\Delta}+\\&(I_{co1}R_{L1}-I_{co2}R_{L2})\end{aligned} \qquad (19)$$

where

$$\Delta=R_{e1}+R_{e2}+\frac{R_{b2}}{\beta_{n2}}\,\frac{R_{b1}}{\beta_{n1}}\,\frac{R_{b1}R_{b2}}{\beta_{n1}\beta_{n2}R_e} \qquad (20)$$

The first term in equation 19 is the desired output for a given input and it represents the voltage gain of the differential amplifier. If the transistors and external resistances are equal, the gain simplifies to

$$\frac{E_{2o}-E_{1o}}{E_{1i}-E_{2i}}=\frac{2\alpha_{n12}R_{L12}}{R_{e12}+\frac{R_{b12}}{\beta_{n12}}+\frac{R_{b}{}^{2}{}_{12}}{2\beta_{n}{}^{2}{}_{12}R_e}} \qquad (21)$$

The second term of equation 19 is the output due to differences in V_{be}. For this to be zero at all temperatures, V_{be1} and V_{be2} must be equal and must drift equally. The third and fifth terms are due to differences in I_{co1} and I_{co2}, and the fourth term is an output due to differences in β_{n1} and β_{n2}.

If R_{e1} and R_{e2} are made into a potentiometer with the movable tap connected to the common emitter resistor, the potentiometer can be adjusted to give zero output for zero input at a particular temperature. That is, the potentiometer can be of such value that it can adjust the fourth term to cancel out the sum of the second, third, and fifth terms. However, if the variation of V_{be} and I_{co} of both transistors does not vary exactly the same with temperature, the difference will be revealed as drift in the output. Since V_{be} and I_{co} variations with temperature are intrinsically a function of the semiconductor, the variations in drift between transistors is considerably less than the drift itself.

To determine the common mode rejection of the differential amplifier, the common mode gain is calculated as follows. Taking only the terms of equation 18 containing input signals, but without making the assumption of $R_{b1}/\beta_{n1}R_e\ll 1$ (since later algebraic addition makes the term important for common mode signals), results in

$$E_{2o}=\frac{1}{\Delta}\left[-\alpha_{n2}R_{L2}\left(\frac{R_{b1}}{\beta_{n1}R_e}E_{2i}+E_{2i}-E_{1i}\right)\right] \qquad (22)$$

Similarly for E_{1o}, and subtracting, yields

$$E_{2o}-E_{1o}=\frac{1}{\Delta}\left[-\alpha_{n2}R_{R2}\left(\frac{R_{b1}}{\beta_{n1}R_e}E_{2i}+E_{2i}-E_{1i}\right)+\alpha_{n1}R_{L1}\left(\frac{R_{b2}}{\beta_{n2}R_e}E_{1i}+E_{1i}-E_{2i}\right)\right] \qquad (23)$$

For a common mode signal $E_{2i}=E_{1i}=E_{cm}$, and the gain is

$$G_{cm}=\frac{E_{2o}-E_{1o}}{E_{cm}}=\frac{1}{\Delta}\times\left(-\alpha_{n2}R_{L2}\frac{R_{b1}}{\beta_{n1}R_e}+\alpha_{n1}R_{L1}\frac{R_{b2}}{\beta_{n2}R_e}\right) \qquad (24)$$

The major effect is due to difference in β; therefore it is assumed that

$$R_{L1}=R_{L2}=R_{L12},\ R_{b1}=R_{b2}=R_{b12},\ \alpha_{n2}=\alpha_{n1}=\alpha_{n12}$$

which simplifies equation 24 to

$$G_{cm}=\frac{1}{\Delta}\left[\frac{\alpha_{n12}R_{L12}R_{b12}}{R_e}\left(\frac{1}{\beta_{n2}}-\frac{1}{\beta_{n1}}\right)\right] \qquad (25)$$

Under the same assumption the differential mode gain in given by equation 21, and the ratio is

$$\mathrm{CMR}=\frac{G_{dm}}{G_{cm}}=\frac{2R_e\beta_{n1}\beta_{n2}}{R_{b12}(\beta_{n1}-\beta_{n2})} \qquad (26)$$

where CMR = common mode rejection.

As might be expected from similar tube amplifiers, the common mode rejection varies directly with the common emitter resistor R_e, and inversely with the differences in β. If identical β transistors were used, the previous assumptions would have to be re-evaluated.

It should be emphasized that the equivalent circuit of Fig. 4(B) assumes an internal collector resistance which is high compared with the collector load resistances; when this is not true, the resulting analysis becomes even lengthier than that shown here.

Analysis of Drift in Differential Amplifiers

Fig. 5(A) illustrates a typical operational amplifier with a differential input. R_{e2} is small in order to keep the impedance at point A small, and R_{e1} is made equal to R_{e2} in order to keep the circuit balanced for drift effects. R_{e1} and R_{e2} are small compared with R_i as is required for operational amplifiers. This results in some of the amplifier input current being shunted into R_{e1} where it produces no

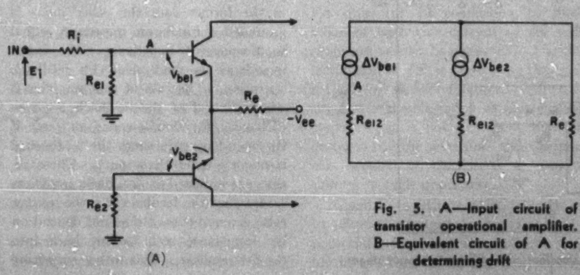


Fig. 5. A—Input circuit of transistor operational amplifier. B—Equivalent circuit of A for determining drift

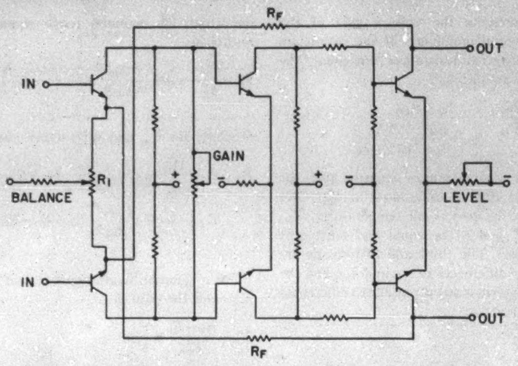


Fig. 6. Differential transistor amplifier with feedback to first-stage emitters

open-loop current gain; however, the gain can be made up in later stages.

The equivalent circuit for drift due to ΔV_{be} is shown in Fig. 5(B), where $R_{e1}=R_{e2}=R_{e12}$, and the internal emitter base impedance is considered small in comparison with R_{e12}. R_t can be neglected or the parallel combination of R_t, and R_{e1} made equal to R_{e12}.

An equal change in V_{be1} and V_{be2} causes current to flow in R_e and the result is attenuated by the common mode rejection of the differential amplifier. By Ohm's law the equivalent drift current at point A due to a difference in the change of V_{be1} and V_{be2} is

$$I_d=\frac{\Delta V_{be1}-\Delta V_{be2}}{R_{e12}+\dfrac{R_{e12}R_e}{R_{e12}+R_e}} \tag{27}$$

To calculate the output voltage due to drift current I_d, the equation for output voltage per input current of an operational amplifier is from equation 48:

$$\frac{E_L}{I_t}=-R_f$$

if

$$Z_f \gg \frac{Z_1}{A_t}$$

and

$$Z_L \gg \frac{Z_f+Z_1}{1+A_t} \tag{28}$$

and

$$E_{L1}=-I_dR_f=-\frac{\Delta V_{be1}-\Delta V_{be2}}{R_{e12}+\dfrac{R_{e12}R_e}{R_{e12}+R_e}}R_f \tag{29}$$

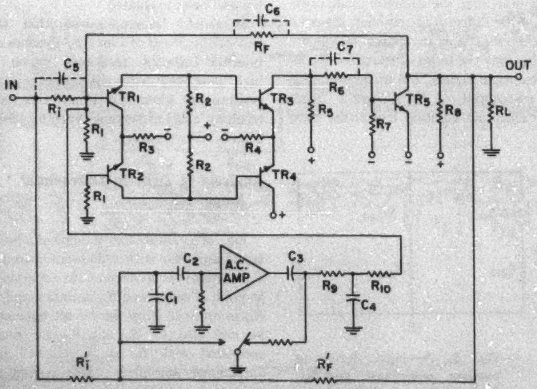


Fig. 7. Chopper stabilized operational transistor amplifier with differential input stages

For unequal I_{co} changes with temperature, the equation is

$$E_{L2}=-\left(\frac{\Delta I_{co1}}{\alpha_1}-\frac{\Delta I_{co2}}{\alpha_2}\right)R_f \tag{30}$$

where E_{L1} and E_{L2} are the output potentials due to V_{be} and I_{co} drifts respectively. These output drifts can be referred to the input by dividing by the closed loop voltage gain given by $-R_f/R_t$, and they become

$$E_{d1}=\frac{\Delta V_{be1}-\Delta_{be2}}{R_{e12}+\dfrac{R_{e12}R_e}{R_{e12}+R_e}}R_t \tag{31}$$

and

$$E_{d2}=\left(\frac{\Delta I_{co1}}{\alpha_1}-\frac{\Delta I_{co2}}{\alpha_2}\right)R_t \tag{32}$$

where E_{d1} and E_{d2} are the voltage drifts due to V_{be} and I_{co} effects respectively, both being referred to the input.

In a typical example, $\Delta V_{be1}-\Delta V_{be2}=17$ maximum millivolts (for a sample of 29 type *2N479*) over a temperature range of 70 to 150 F, $R_{e12}=10$ kilohms, $R_e=29$ kilohms, $R_t=100$ kilohms, $E_{d1}\doteq 100$ millivolts. This can be reduced by chopper stabilization to 0.5 millivolt if drift reduction is 200; see equation 10.

For silicon transistors type *2N479*, $(\Delta I_{co1}-\Delta I_{co2})=0.01$ microampere maximum (for a sample of ten units), and this corresponds to an $E_{d2}=1$ millivolt, which is negligible when reduced by a factor of 200. However, the same theory will result in low drifts if germanium differential stages are used.

Practical Differential Amplifiers

The Slaughter amplifier is a differential amplifier with feedback, and it incorporates all the advantages of this type of operation. However, it requires that the output be single-ended and that it be in phase with the input. Fig. 6 illustrates a differential amplifier with feedback that allows for a single-ended output of either phase, or a differential output. The input can be single- or double-ended; in the former case the other input is grounded through an impedance equal to the source impedance. The input impedance is made high by feedback operation. The gain of this amplifier is the reciprocal of the feedback β ($A_v=(2R_f+R_1)/R_1$, double-ended output) if the open-loop gain times the feedback β is much greater than unity. These results are standard for feedback amplifiers except that the feedback β is the resistor ratio given above; it does not depend on the impedance seen looking back into the first emitter. This rather surprising result has been confirmed by analysis. Incidentally, the analysis applies to a single-ended amplifier in which the feedback is introduced at the first-stage emitter in grounded emitter operation.

An amplifier was constructed according to Fig. 6 and performed as expected. The balance, gain, and level controls behave as they would in an equivalent tube-type amplifier. The gain control is optional but it is useful practically to stop oscillations if the open-loop gain is too high for the phase shifts introduced by the transistors. A gain of 10 was achieved with 30-db (decibel) feedback without frequency compensation with *2N479* silicon transistors. The divider network between the second and third stage is required to bring the output to zero d-c volts and is an ideal place to introduce frequency compensation if the transistor phase shifts cause oscillations. (The details of frequency stabilization are discussed in a later section.) The drift was, of course, a function of the difference in drift characteristics of the first-stage transistors.

To take advantage of chopper stabilization and differential stage drift reduction, an operational amplifier can be used with a differential input and a single-ended output. The chopper stabilization is effective, however, only for a single-ended input signal; Fig. 7 illustrates this type of amplifier. The gain is given by the ratio of R_f to R_i, the input impedance is the parallel combination of R_1 and R_i', and the drift reduction is equal to R_i'/R_i times the net current gain of the a-c amplifier.

The a-c amplifier has a net phase reversal so that the single-pole double-throw chopping yields no net phase reversal for the direct current. It is possible to feed the output of the chopper amplifier into the base of TR_2 if the a-c amplifier and chopper have a net d-c phase reversal.

The important features in the design of this type of amplifier are summarized as follows:

D-C DESIGN

The first and second stages are the differential type, and to determine their operating points the collector current is first specified. Since the emitter potentials are very close to the base, the common emitter resistor is calculated to have twice the individual collector currents flowing when it is operating between the base potential and the negative supply.

R_3 and R_4 are chosen high for good common mode rejection (see equation 26 for the common mode rejection formula) and this determines their negative return potential.

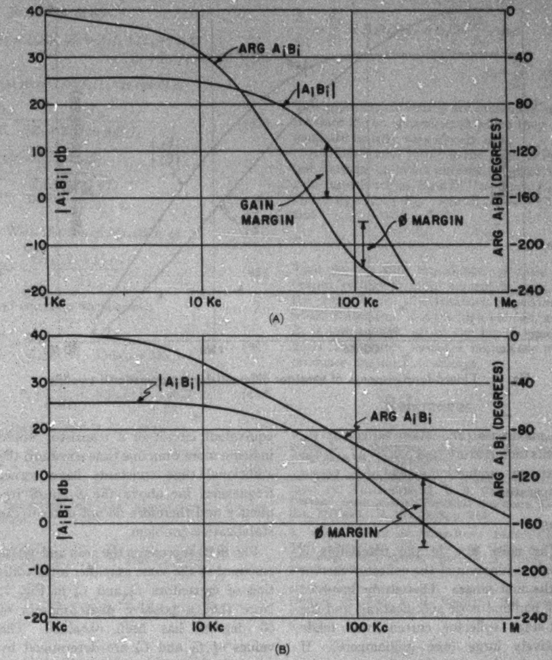


Fig. 8. A—Open-loop gain and phase response of transistor differential input operational amplifier with no frequency compensation. B—After stabilization to self-oscillation

The collector load resistance can then be selected so that it will be high (perhaps ten times as high) compared to the input impedance of the following stage. This is required since, ideally, all of the signal current should flow into the next stage. The collector supply can then be chosen large enough to allow for the signal swing, with the low value of a-c load of the next stage taken into account. For quiescent d-c conditions the small value of base bias current of the next stage can usually be neglected. The collector load of TR_4 is omitted so that it is actually an emitter follower. Since this stage is a double-ended to single-ended converter, it is not necessary, and the design is still the same since the collectors can usually be considered as current sources of high impedance.

R_6 and R_7 constitute a divider network to allow the output to be at zero d-c potential. Knowing the collector potential of TR_3 and realizing that the base of TR_5 is essentially at emitter supply potential, lead to the choice of R_5, R_6, and R_7, so that the base current of TR_5 is at the desired quiescent value. R_6 and R_5 divide the signal current so as to reduce the gain, and it is therefore desirable to keep R_6 at a minimum. However, this demands a maximum of bleeder current through R_5, R_6, and R_7 and a compromise is necessary. R_7 is made much higher than the input impedance of TR_5 in order not to divert a lot of signal current, and this usually requires its negative return supply to be rather high negatively.

It is possible to use an opposite semiconductor material for the middle stage and eliminate the divider loss; however, R_6 is an ideal location to shunt a capacitor to form a lead network for frequency stabilization. This will be discussed later.

DRIFT DESIGN

Equations 31 and 32 yield the drift referred to the input where $R_{e12}=R_1$, R_i is the same, $R_i>>R_1$, $\Delta V_{be1}-\Delta V_{be2}$ is

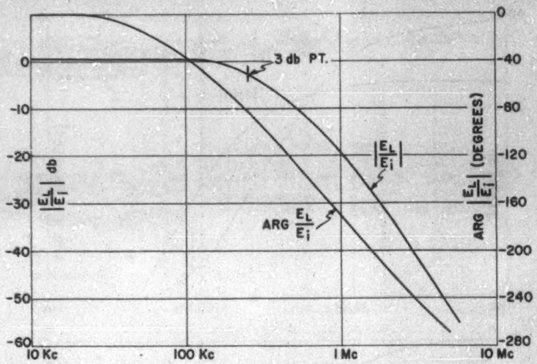


Fig. 9. Closed-loop response of transistor differential input operational amplifier

the maximum production spread of V_{be} versus temperature, and $\Delta I_{co1}-\Delta I_{co2}$ is the maximum production spread of I_{co} versus temperature.

Noise

The noise due to the transistors is mainly a function of the collector current in the first stage. This study involved very nominal noise specifications and the first-stage collector current was made relatively large (one milliampere). If this value is reduced to decrease noise effects, care must be taken to ensure that the transistor does not drift out of its active region with temperature. The small common mode gain of the differential amplifier is a distinct advantage in this respect compared with a single-ended type of amplifier.

A-C Design

If a large dynamic range is required, the d-c design must allow for the specified swing without saturating or cutting off. Again, the differential amplifier, due to its low common mode gain, helps maintain the room temperature operating conditions when the temperature varies.

To a first approximation, each transistor stage introduces a 6 db per octave slope with a corner frequency equal to the β cutoff frequency. Fig. 8(A) is a typical open-loop gain and phase response for a 3-stage amplifier shown in Fig. 7. Since there are three time constants involved, self-oscillations are possible, and indeed Fig. 7 indicates a negative gain and phase margin. Note that the gain drops off at a rate of 18 db per octave, indicating that the three β cutoff corner frequencies are close to each other. Note also that although the high-frequency equivalent circuit of a transistor would indicate more than one time constant, the additional time constants have corner frequencies far above the β cutoff frequency and therefore do not enter in the stabilization problem.

Fig. 8(B) represents the gain and phase response of the same amplifier after addition of capacitors C_2 and C_8 in Fig. 7. Note that a positive phase margin of 60 degrees has been obtained. The values of C_7 and C_8 are determined by standard techniques and will not be discussed here. There are many stabilization networks available; the two chosen are typical. Note that the use of the bleeder network R_5, R_6, and R_7 allow C_2 to form a simple cascade lead network.

Fig. 9 represents the closed loop (output divided by input) of the amplifier in Fig. 8(A) when the open-loop gain and phase were adjusted as in Fig. 8(B). The half-power point is at 250 kc. This could be raised by adjusting C_2 and C_8 in order to have a peaking in the closed loop response, but Fig. 9 represents the maximum response without peaking. Of course, peaking would reduce the phase margin of 60 degrees, but this is more than adequate to ensure against self-oscillation.

The capacitor C_8 of Fig. 7, can be used to shape the closed loop response without affecting the open-loop response of the amplifier. The reason for this is that the low input impedance at the base of TR_1 effectively short-circuits R_1 and C_8. The combination of R_1, C_8, and the input impedance of TR_1 form a lead network which which will introduce a rising 6 db per octave slope in the closed loop response, permitting a greater bandwidth than shown in Fig. 9.

Chopper Amplifier and Filter

The a-c amplifier shown in Fig. 7 represents a fairly simple design since it must amplify only the chopper frequency. A 3-stage common emitter germanium transistor amplifier was used with large emitter resistances to ensure a low stability factor. These resistors are heavily by-passed so they do not reduce the gain at the chopper frequency. The combination of net phase reversal in the chopper amplifier and the single-pole double-throw chopping contacts yield a zero net phase shift for d-c signals. The d-c output is fed back to the base of the first transistor. If other combinations of a-c amplifier phase reversal and chopping contacts yield a net d-c phase reversal, the output can be fed into the base of TR_3 in Fig. 7.

C_1 together with R_4' form a filter to prevent chopping frequency signals to be fed back to the input, and C_2 and C_3 are blocking capacitors. R_9, C_4, and R_{10} are the output filter. To reduce the chopping frequency ripple to a negligible amount, the band pass of the entire chopper amplifier is reduced to a few cps (cycles per second). This is sufficient for correcting d-c drifts but it should be noted that a more complex filter will provide additional advantages. If the filter were designed to have a flat band pass close to the chopping frequency and the fall rapidly to give the desired attenuation at the chopping frequency, stabilization could be obtained over a reasonably wide bandwidth, one example might be to use a 400-cps chopping frequency, and a filter which was flat out to > 120 cps. This combination would allow simple power supplies to be used since the 60- to 120- cps ripple introduced by such supplies would be attenuated sharply by the chopper stabilization technique. The analysis of such stabilization would be similar to that already shown.

Appendix

The current operational amplifier of Fig. 2(B) can be analyzed assuming $-A_i$ is the open-loop current gain which in general is a function of Z_L and Z_f.

From Fig. 2(B):

$$I_1=I_i+I_f=\frac{I_L+I_f}{-A_i} \tag{33}$$

Define

$$\beta_i=\frac{I_f}{I_R} \tag{34}$$

Equation 33 becomes

$$I_i+\beta_i I_L=\frac{I_L+\beta_i I_L}{-A_i}$$

$$\frac{I_L}{I_t}=\frac{-A_t}{(1+\beta_t+A_t\beta_t)}=\frac{-A_t}{1+\beta_t(1+A_t)} \tag{35}$$

From Fig. 2(B):

$$E_L=I_LZ_L=I_fZ_f+(I_t+I_f)Z_1 \tag{36}$$

where Z_1 is the input impedance of the current amplifier; from equation 33

$$I_t=-\frac{I_f(A_t+1)+I_L}{A_t}$$

Putting this into equation 36 results in

$$\frac{I_f}{I_L}\equiv\beta_t=\frac{Z_LA_t+Z_1}{A_tZ_f-(A_t+1)Z_1+Z_1A_t}$$

Dividing by A_t, and simplifying yields

$$\frac{I_f}{I_L}\equiv\beta_t=\frac{Z_L+\dfrac{Z_1}{A_t}}{Z_f-\dfrac{Z_1}{A_t}} \tag{37}$$

If, as is usually the case,

$$Z_L\gg\frac{Z_1}{A_t};\ Z_f\gg\frac{Z_1}{A_t}$$

$$\beta_t\doteq\frac{Z_L}{Z_f} \tag{38}$$

From equations 35 and 37

$$\frac{I_L}{I_f}=-\frac{(A_tZ_f-Z_1)}{Z_f+Z_1+Z_L(1+A_t)} \tag{39}$$

$$\doteq-\frac{Z_f}{Z_L}\text{ if }Z_f\gg\frac{Z_1}{A_t},\ Z_L\gg\frac{Z_f+Z_1}{1+A_t} \tag{40}$$

From Fig. 2(B)

$$E_L=I_LZ_L,\ E_t=Z_tI_t+(I_t+I_f)Z_1 \\ =(Z_t+Z_1)I_t+Z_1I_f \tag{41}$$

and since $I_f\equiv\beta_tI_L$, the result is

$$\frac{E_L}{E_t}=\frac{Z_LI_L}{(Z_t+Z_1)I_t+\beta_tZ_1I_L} \tag{42}$$

from equation 33

$$I_t=-\frac{I_f(A_t+1)+I_L}{A_t}$$

With the use of equation 34

$$I_t=-\frac{\beta_tI_L(A_t+1)+I_L}{A_t} \tag{43}$$

and equation 43 becomes

$$\frac{E_L}{E_t}=\frac{-A_tZ_L}{Z_t[\beta_t(A_t+1)+1]+2Z_t} \tag{44}$$

$$\doteq-\frac{Z_L}{Z_t\beta_t}\text{ if }A_t\beta_t\gg1,\ A_t\beta_t\gg\frac{2Z_t}{Z_t} \tag{45}$$

If, further, $\beta_t\doteq Z_L/Z_f$, then

$$\frac{E_L}{E_t}\doteq\frac{Z_f}{Z_s} \tag{46}$$

Summing up the approximate results yields

$$\frac{I_L}{I_t}\doteq-\frac{Z_f}{Z_L}\text{ if }Z_f\gg\frac{Z_1}{A_t}\text{ and }Z_L\gg\frac{Z_f+Z_1}{1+A_t}$$

$$\beta_t\equiv\frac{I_f}{I_L}\doteq\frac{Z_L}{Z_f}\text{ if }Z_L\gg\frac{Z_1}{A_t}\text{ and }Z_f\gg\frac{Z_1}{A_t}$$

$$\frac{E_L}{E_t}\doteq\frac{Z_f}{Z_t}\text{ if }A_t\beta_t\gg1,\ A_t\beta_t\gg\frac{2Z_1}{Z_t},\text{ and}$$

$$\beta_t\doteq\frac{Z_L}{Z_f} \tag{47}$$

The approximate results are extremely useful since the conditions upon which they depend are usually very closely met.

Another useful result is the ratio of output voltage to input current. Define the gain impedance as $Z_g\equiv E_L/I_t$. Since $E_L=I_LZ_L$, $Z_g=I_L/I_t\ Z_L$. But $I_L/I_t\doteq-Z_f/Z_L$ and

$$Z_g\doteq-Z_f \tag{48}$$

That is, the gain impedance, or ratio of output voltage to input current is equal to the negative of the feedback impedance, which states that if A_t is high enough, and Z_1 is low enough, all of the input current flows through the feedback impedance and produces the output voltage.

References

1. FEEDBACK STABILIZED TRANSISTOR AMPLIFIERS, D. W. Slaughter. *Electronics*, New York, N. Y., vol. 28, 1955, p. 174.

2. SOME TRANSISTOR INPUT STAGES FOR HIGH GAIN DC AMPLIFIERS, G. B. B. Chaplin, A. R. Owens. Institution of Electrical Engineers, London, England, vol. 105, 1958, p. 249.

3. STABILIZATION OF WIDE-BAND DIRECT-CURRENT AMPLIFIERS FOR ZERO AND GAIN, A. W. Goldberg. *RCA Review*, Princeton, N. J., vol. 11, 1950, p. 296.

4. TRANSISTOR CIRCUITS FOR ANALOG AND DIGITAL SYSTEMS, F. H. Blocher. *Bell System Technical Journal*, New York, N. Y., vol. 35, 1956, p. 295.

Instabilities of Push-Pull Magnetic Amplifiers Feeding the Field of an Electric Machine

H. F. STORM
MEMBER AIEE

THE CAUSES of instability of inductively loaded, center-tap magnetic amplifiers, and the remedies for this instability are well known.[1-4] If such a stabilized magnetic amplifier is used to energize the field of an electric machine, such as the shunt field of a d-c generator, the armature voltage of this machine will increase or decrease smoothly with the control voltage of the magnetic amplifier. Since the load current of the magnetic amplifier is unidirectional, the armature voltage of the d-c generator will also be unidirectional. However, in many closed-loop control systems it is necessary to obtain duodirectional output of the d-c geneator. In this case the d-c generator will be equipped with two shunt fields; each shunt field is connected to a magnetic amplifier with such polarity that the magnetomotive forces produced by the two shunt fields are subtractive. Then, if the first magnetic amplifier is controlled "full on" and the second magnetic amplifier is controlled "full off," the armature of the d-c generator develops a voltage of one polarity, and if the control to the magnetic amplifiers is reversed, the armature voltage of the d-c generator is also reversed. One may expect, furthermore, that if each magnetic amplifier is turned "one half on," the net magnetomotive force will be zero, and, hence, that the armature voltage of the d-c generator will also be zero; one may also expect that for a gradual increase of control current of one amplifier, and a similar gradual decrease of control current of the other amplifier, the armature voltage of the d-c generator will show a gradual change. The last two expectations, however, are not likely to be fulfilled. Instead, a minute change of the magnetic amplifier control currents will cause a sudden jump of armature voltage from perhaps +50% to −50% (or vice versa) of rated d-c generator armature voltage, and intermediate values of armature voltage will not be obtainable. This instability renders almost useless this otherwise inviting control scheme. It is suspected

---

Paper 56-730, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer and Pacific General Meeting, San Francisco, Calif., June 25–29, 1956. Manuscript submitted April 12, 1956; made available for printing October 14, 1959.

H. F. STORM is with the General Electric Company, Schenectady, N. Y.

The author wishes to thank Manuel Ares for the experimental information contained in this paper.

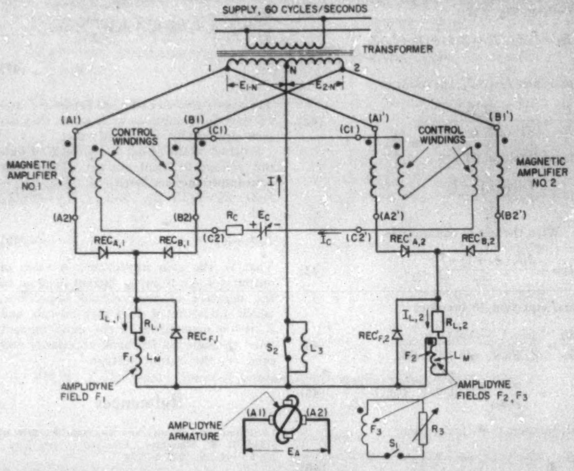


Fig. 1. Magnetic amplifiers 1 and 2, providing push-pull control for amplidyne control fields F_1 and F_2

immediately that magnetic amplifiers are prone to produce instabilities when feeding inductive loads, such as the field of the d-c generator. However, as soon as one of the two magnetic amplifiers is disconnected from its generator field, the other will produce perfect (unidirectional) control of the d-c generator, and vice versa. Hence the instability is of a different nature than suspected.

Origin of the Instability

The circuit under study is shown in Fig. 1. The d-c generator to be controlled by magnetic amplifiers 1 and 2 is of the amplidyne type. The amplidyne control fields F_1, and F_2 are connected to the magnetic amplifiers with opposite polarities, as previously described. Two rectifiers, $REC_{F,1}$ and $REC_{F,2}$ are connected across the fields of the amplidyne in accordance with standard practice, in order to remove the instabilities of the individual magnetic amplifiers when they are feeding inductive loads.

The control windings of the magnetic amplifiers are connected with opposing polarities; thus a change in the common control current I_C will increase the load current of one magnetic amplifier and decrease the load current of the other. The bias windings of the magnetic amplifiers (not shown in Fig. 1) may be energized by a constant direct current, whose value has no first-order significance for the problem under discussion; more information regarding the magnetic amplifiers and the amplidyne are found in the Appendix.

By varying the magnitude and polarity of the magnetic amplifier control voltage E_C (Fig. 1), the magnetic amplifier control current I_C is varied, and the amplidyne armature voltage E_A is expected to conform to the typical push-pull characteristic of Fig. 2(A). Instead, one will discover with considerable dismay that the actual characteristic, shown in Fig. 2(B), is quite different. If the amplifier system is operating at point A, Fig. 2(B), and the control current I_C is decreased, the amplidyne armature voltage E will decrease slowly until point B is reached. Any further decrease in amplistat control current I_C causes the amplidyne armature voltage E_A to snap from the positive value at B to the negative value at C. Similarly, if the control current I_C is increased, starting from D, the armature voltage E_A will vary along the curve D-C until point E is reached, at which point snap action will again occur. Experiments indicate that each amplistat is, by itself, well behaved, and hence one concludes that the instability, called snap action, is caused by the interaction of the two magnetic amplifiers.

Interaction between the two magnetic amplifiers can occur along two channels: 1. the common control circuit, and 2. the common amplidyne control flux. During steady-state operation, interaction through the two channels can be accomplished only by a-c components; experimentation supports the theory that those a-c components contain only harmonics of even order. It is beyond the scope of this paper, however, to discuss in greater detail the mechanics of even-harmonic generation.

Remedies for Instabilities

The evidence thus far discussed leads to the following remedies.

1. Increase of Control-Circuit External Impedance

Increasing the external impedance of the control circuit may be achieved by

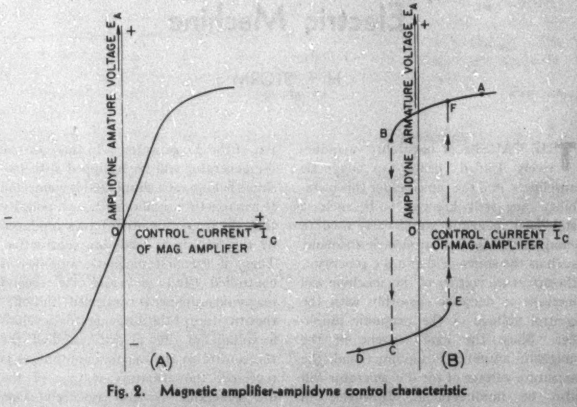


Fig. 2. Magnetic amplifier-amplidyne control characteristic

A—Expected control characteristic
B—Actual control characteristic, exhibiting instability, or snap action

Vol.
Pt. 1
196

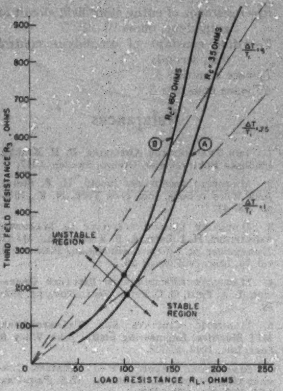


Fig. 3. Relation between R_L and R_3 to eliminate instability, and per-unit increase $\Delta T/T_1$ of time constant of amplidyne control flux

increasing the external control circuit resistance R_C, or by inserting a reactor into the external control circuit.

Two experimental characteristics which form the boundary between stable and unstable operation are shown in Fig. 3. One characteristic has been taken with a total control circuit resistance $R_C=35$ ohms, the other with $R_C=160$ ohms. The space between the curves is indicative for unstable operation with $R_C=35$ ohms, but stable operation with $R_C=160$ ohms.

2. Transformation Attenuation

To attenuate the transformation of alternating voltages from one amplidyne control field to the other, several methods are suggested which will be explained by means of the equivalent circuit diagram for a-c components in the amplidyne fields; see Fig. 4. This diagram has been derived from the equivalent transformer circuit; it neglects leakage reactances and core loss.[5] It will be seen that interaction can be combatted by the following means:

Increasing Load Resistances $R_{L,1}$, $R_{L,2}$ of Magnetic Amplifiers. For a fixed control circuit resistance R_C, and a fixed third field resistance R_3, discussed below, an increase of load resistance R_L leads in the direction of increased stability, as evidenced in Fig. 3. It should be pointed out, however, that with a given capacity of the magnetic amplifiers there is a limit as to the maximum amount of load resistance that can be afforded.

Inserting Reactors in Series with Load Resistors $R_{L,1}$, $R_{L,2}$. Some further thought will indicate that the same effect can be obtained by a single reactor L_3 in the lead to the neutral N, with switch S_2 open; see Fig. 1. The physical size of such a reactor is usually quite small. A reactor of a type size of only $2\pi f L_3 I^2 =$ 200 volt-amperes will effectively reduce the snap action tendency of a large 25-kw output unit.

Increasing Admittance Across Terminals of Magnetizing Inductance $L_{L,1}=L_{L,2}=L_M$. The most obvious solution is to connect resistors in parallel to the amplidyne fields F_1 and F_2. While this method is effective in removing the instability, the additional power consumed by these resistors calls for an undesirable increase in magnetic-amplifier capacity. The same salutary effect, however, can be achieved without an increase in magnetic-amplifier capacity by using a third field F_3 of the amplidyne.[6] This field F_3 is connected to an external resistor R_3 by switch S_1, which is now considered as closed; the power dissipated in resistor R_3 is usually only a fraction of one watt; see Fig. 1.

The effectiveness of the last-mentioned method can be gaged by an inspection of Fig. 3; the degree of amplidyne armature loading has no measurable effect on the stability of the control. Stable operation of the amplifier system is obtained for the regions to the right of and below the curves. One might conclude from curves A and B that the best design will result if R_3 is made as low as possible; this however is not necessarily so, for the following reason.

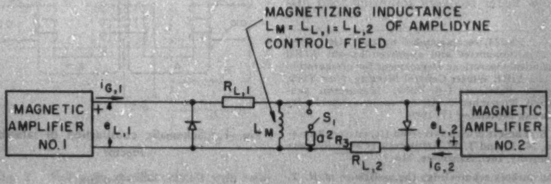


Fig. 4. Simplified equivalent a-c circuit of amplidyne fields

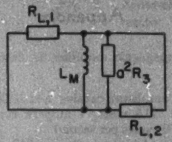


Fig. 5. Equivalent circuit for determining the time constant of amplidyne control flux

The flux time constant of a field having several windings terminated into purely resistive circuits, is equal to the sum of the time constants of the several windings.[7] Hence, the time constant T_C of the amplidyne flux is from Fig. 5:

$$T_C=\frac{L_M}{R_{L,1}}+\frac{L_M}{R_{L,2}}+\frac{L_M}{a^2R_3} \tag{1}$$

For

$$R_{L,1}=R_{L,2}=R_L \tag{2}$$

one obtains

$$T_C=T_1+\Delta T \tag{3}$$

where T_1 is

$$T_1=\frac{L_M}{R_L/2} \tag{4}$$

and ΔT is the time constant which is contributed by the presence of a finite R_3

$$\Delta T=\frac{L_M}{a^2R_3} \tag{5}$$

The per-unit increase of the amplidyne time constant $\Delta T/T_1$, due to the presence of R_3, follows from equations 4 and 5:

$$\frac{\Delta T}{T_1}=\frac{R_L/2}{a^2R_3} \tag{6}$$

One concludes that, while an increase of R_L/R_3 has the agreeable effect of counteracting the instability shown in Fig. 3, it also has the disagreeable effect of increasing the time constant T_C of the amplidyne control field. To facilitate the proper selection of R_L/R_3, the loci for various ratios of $\Delta T/T_1$ have been added to Fig. 3.

Conclusions

The various remedies for instability, listed in the previous paragraphs, are often used most effectively in combination. After the instability has been removed, the system functions according to the expectations illustrated in Fig. 2 (A). Many such magnetic-amplifier-amplidyne combinations are in use in industry and in utility fields. While this paper presents several methods of combating instability, it should not be inferred that no other methods exist.

Appendix

The magnetic amplifier is described as follows:

Type: amplistat catalog no. *101D511-G1*.
Rating: 115 volts, 1 ampere continuous, 60 cycles per second.
Control and bias windings: 250 turns, 4.9 ohms.
Gate winding: 1,000 turns, 9 ohms.
Core material: grain-oriented silicon-iron.

The amplidyne is described as follows:

Serial no. *5AM79A8324*.
Rating: 250 volts, 1.5 kw, 1,725 rpm.
Control fields F_1 and F_2: 205 turns per pole, 45 ohms.
Third field F_3: 400 turns per pole, 55 ohms.
Fourth field (not used): 330 turns per pole, 116 ohms.
Control field inductance for 1,000 turns per pole: 50 henrys.
Control ampere-turns per pole for 200-volt armature voltage: 40.
Control ampere-turns per pole for 250-volt armature voltage: 80.

Nomenclature

$a = N_C/N_3$
E_C = control voltage for magnetic amplifiers, volts, average
E_A = amplidyne armature voltage, volts, average
f = frequency of supply voltage
I = current in neutral of supply transformer, amperes, average
I_C = control current for magnetic amplifiers, amperes, average
L_3 = inductance, henrys
N_C = number of turns of amplidyne control winding F_1 or F_2
N_3 = number of turns of third amplidyne control winding F_3
L_M = magnetizing inductance of amplidyne control field F_1 or F_2, henrys
R_C = control circuit resistance of amplistat circuit, ohms
R_L = load resistance of magnetic amplifier, including amplidyne field resistance, ohms
R_3 = resistance of entire third field circuit of amplidyne, ohms
T_C = time constant of amplidyne control flux, seconds
T_1 = see equation 4
ΔT = see equation 5

References

1. The Transductor Amplifier, U. H. Krabbe. Lindhska Boktryckeriet, Orebro, Sweden, 1947.

2. Magnetic Amplifiers (book), H. F. Storm. John Wiley & Sons, Inc., New York, N. Y., 1955, pp. 320–41.

3. Inductive Load Instability in Magnetic Amplifiers, H. L. Leon, A. B. Rosenstein. *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 74, Mar. 1955, pp. 8–20.

4. Magnetic Amplifiers in Bistable Operation, L. A. Finzi, G. C. Feth. *Ibid.*, Nov., pp. 592–98.

5. Magnetic Circuits and Transformers, MIT Electrical Engineering Staff. John Wiley & Sons, Inc., 1943, p. 338.

6. Stabilizing Circuit for Alternating Current Generators, H. F. Storm. *U.S. Patent no. 2,724,797*, Nov. 22, 1955.

7. See reference 2, p. 161.

Saturable-Reactor Control of Full-Wave and Biphase Rectifiers

J. B. THOMAS
MEMBER AIEE

J. W. DRENNING
MEMBER AIEE

IN MANY rectifier applications means must be provided to vary the output voltage. In some cases discrete control is permissible while in others the output must be continuously variable. A common method of providing the latter is to employ a saturable-core reactor in series with the primary of the rectifier transformer.

This paper treats certain problems peculiar to the design of rectifiers with saturable-reactor controls. One of the principal factors to be considered is the abnormally high voltage which may exist at various points in the circuit. This condition may become particularly acute in biphase and half-wave connections.

Operation of Saturable-Core Reactors

The impedance of a saturable-core reactor is made variable by means of a control winding in which a direct current flows and determines the saturation level in the iron core. As a rectifier control, this device is conventionally connected in series with the a-c lines supplying the rectifier transformer. Voltage supplied to the rectifier is therefore determined by the instantaneous difference between line and saturable-reactor voltages, the latter being variable by means of the control winding previously mentioned.

A saturable-core reactor is shown schematically in Fig. 1. The a-c coils of the reactor may be connected in parallel or in series.[1,2] Most of the "stock" saturable reactors are parallel-connected since the a-c components induced in the d-c winding of the series-connected reactor are frequently objectionable. However, series-connected reactors have an inherently faster response time; hence they are used in such applications as electrostatic precipitator power supplies to allow faster recovery following load disturbances. For brevity, only the series-connected reactor will be considered in the following discussion, although essentially the same problems exist with the parallel-connected reactor.

In the series-connected reactor (terminals *2* and *3* connected in Fig. 1), the same current flows through each a-c coil. If this current is given by

$$i_n = I_n \sin (n\omega t + \phi_n) \tag{1}$$

then reactor voltage e_{14} may be expressed as a sum of the fundamental and odd harmonics:

$$e_{14} = e_n = \sum_k E_{kn} \sin (kn\omega t + \phi_{kn}) \qquad k = 1, 3, 5, \ldots \tag{2}$$

Operation of the reactor cores over the nonlinear portion of the magnetization curve generates the harmonics and symmetry dictates that they be odd.

The relationships in equations 1 and 2 are shown graphically in Fig. 2 for a given d-c saturation level chosen so that the first and third harmonics predominate in

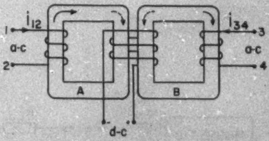


Fig. 1. Schematic construction of saturable reactor

---D-c flux
——A-c flux

Paper 59-223, recommended by the AIEE Electronics Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 1–6, 1959. Manuscript submitted November 3, 1958; made available for printing September 17, 1959.

J. B. Thomas is with Princeton University, Princeton, N. J., and J. W. Drenning is with Koppers Company, Inc., Baltimore, Md.

The authors acknowledge the assistance of H. T. Williams of Koppers Company, Inc., in preparing this paper.

Vol.
Pt.
196

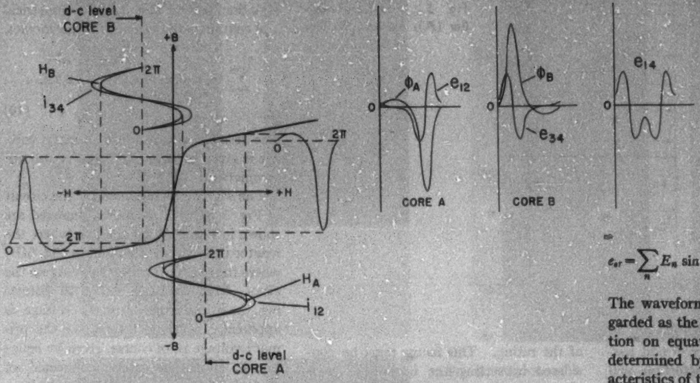


Fig. 2. Series-connected reactor

the reactor voltage e_{14}. This particular case was selected because it illustrates the actual waveforms encountered in the reactors used in this study and discussed later. In Fig. 2 the phase shift due to hysteresis has been ignored in the interest of simplicity.

As far as this discussion is concerned two cases may be distinguished. The first case is the full-wave rectifier where the reactor current is mirror-symmetrical (henceforth called symmetrical) and therefore contains only odd harmonics.[3] The second case is the unsymmetrically loaded biphase rectifier or the half-wave rectifier where the reactor current is unsymmetrical and therefore contains even and odd harmonics.[3]

Full-Wave Rectifier

Fig. 3 shows a schematic diagram of a full-wave bridge rectifier using a series-connected saturable reactor as the primary control device. An *R-C* load is shown since this is frequently encountered in practice and poses a more difficult problem than a purely resistive load. The transformer *T* isolates the load from the line or provides a change in voltage or both.

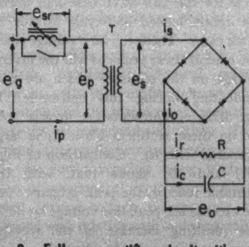


Fig. 3. Full-wave rectifier circuit with saturable-reactor control

As previously stated the reactor current i_p is symmetrical since the secondary current i_s is symmetrical. In general, therefore, the reactor current i_p, although not sinusoidal, can be expressed as the Fourier series:[3]

$$i_p=\sum_n I_n \sin(n\omega t+\phi_n) \quad n=1,3,5,\ldots \qquad (3)$$

As shown in equations 1 and 2, the voltage drop across the reactor for a sinusoidal current flow therein can be expressed as a sum of the fundamental and odd harmonics. Although the reactor is not linear and superposition does not hold, the net voltage drop across it due to the current of equation 3 is still a sum of the fundamental and odd harmonics, i.e.,

$$e_{sr}=\sum_n E_n \sin(n\omega t+\theta_n) \quad n=1,3,5,\ldots \qquad (4)$$

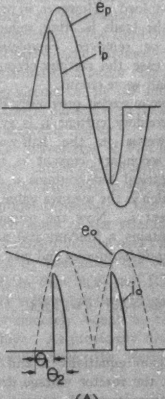
The waveform of equation 4 can be regarded as the result of a nonlinear operation on equation 3, the operator being determined by the magnetization characteristics of the saturable reactor and by the d-c saturation level. Consequently the reactor drop e_{sr} will still be symmetrical but the relative amount of the higher order harmonics will have been increased.

The resulting transformer primary voltage will be:

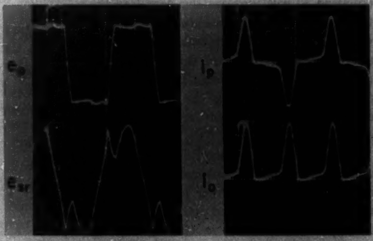
$$e_p=e_g-e_{sr} \qquad (5)$$

Thus the presence of the reactor has created a nonsinusoidal transformer-exciting voltage which is, however, symmetrical.

The effects just discussed are illustrated and compared in Figs. 4(A) and (B). In Fig. 4(A) the waveforms for the full-wave rectifier of Fig. 3 are shown for the case where the saturable reactor is short-circuited and hence the transformer exciting voltage is sinusoidal.[4] In Fig. 4(B) the waveforms for the circuit of Fig. 3 are shown for the case where the saturable reactor is not short-circuited. The effect of the reactor in distorting the transformer primary voltage is easily seen. The waveform is almost rectan-



(A)



(B)

Fig. 4. Waveforms in a full-wave rectifier

A—R-C load
B—Saturable-reactor control

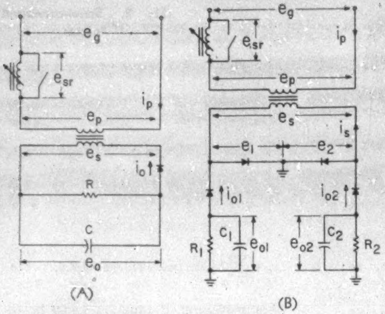


Fig. 5. Half-wave rectifier (A), biphase rectifier (B)

gular indicating the presence of odd harmonics. The degree of distortion will vary, of course, depending on the circuit parameters, but some typical values of transformer primary voltage are given in the following. Without reactor:

$$e_p = e_g = E_g \sin \omega t \tag{6}$$

With reactor:

$$e_p = E_{p1} \sin (\omega t + \psi_1) + 0.39 E_{p1} \sin (3\omega t + \psi_3) + 0.22 E_{p1} \sin (5\omega t + \psi_5) + 0.07 E_{p1} \sin (7\omega t + \psi_7) + \dots \tag{7}$$

Equation 7 was obtained experimentally with a wave analyzer using narrow-band filters.

Fig. 4(B) also shows the waveform of the saturable-reactor voltage. This waveform exhibits the characteristic "double hump" distortion shown in Fig. 2. Analysis of this voltage yields the following expression:

$$e_{sr} = E_{sr1} \sin (\omega t + \theta_1) + 0.33 E_{sr1} \sin (3\omega t + \theta_3) + 0.21 E_{sr1} \sin (5\omega t + \theta_5) + 0.17 E_{sr1} \sin (7\omega t + \theta_7) + 0.07 E_{sr1} \sin (9\omega t + \theta_9) + 0.06 E_{sr1} \sin (11\omega t + \theta_{11}) + \dots \tag{8}$$

The presence of the harmonics in the transformer primary voltage, as shown in equation 7, will increase the core loss in the transformer. In addition the peak value of this voltage will be less than that of a sinusoidal voltage of the same effective value. This fact must be taken into account in determining the transformer turns ratio required for a given peak output voltage.

The reactor, even when fully saturated, will prevent full-line voltage from being applied to the transformer primary, particularly at high load currents. The transformer, therefore, must be rated at some lower voltage. However, should the secondary be open-circuited or the load appreciably reduced, the reactor drop may not be sufficient to prevent transformer primary voltages in excess of the rating. This factor must be considered in setting the insulation levels of the transformer and the inverse voltage rating of the rectifier elements. It is particularly important in high-voltage designs.

Biphase and Half-Wave Rectifiers

Two types of half-wave rectifier circuits are shown in Fig. 5. The conventional half-wave circuit of Fig. 5(A) is often employed in electronic systems where current requirements are low and simplicity is important. The circuit of Fig. 5(B) is actually a special case of the full-wave rectifier and is often called a biphase or double-half-wave circuit. This circuit is well known in the electostatic precipitator industry where it is used to obtain the advantages of half-wave energization without the disadvantages of the half-wave circuit. Two separate half-wave *R-C* loads are supplied alternately every one-half cycle, thereby causing the rectifier to operate as though it were supplying a single full-wave load. However, in practice, one of the loads is often reduced substantially or removed completely. In the latter case the rectifier becomes essentially a half-wave circuit.

In the full-wave circuit previously discussed the secondary current i_s is symmetrical. However in the half-wave rectifier, the secondary current is not only unsymmetrical but contains a d-c component which is the average value of the load current i_o. Now the primary current i_p contains a fundamental and both even and odd harmonics and may be written as[4]

$$i_p = \sum_n I_n \sin (n\omega t + \phi_n) \quad n = 1,2,3,4,\dots \tag{9}$$

Considering again equations 1 and 2, it follows that the reactor voltage drop due to the unsymmetrical current of equation 9 will also be unsymmetrical and contain both even and odd harmonics, i.e.,

$$e_{sr} = \sum_n E_n \sin (n\omega t + \theta_n) \quad n = 1,2,3,4,\dots \tag{10}$$

The resulting transformer primary voltage is given by equation 5 and is also unsymmetrical.

The waveforms obtained for the circuit of Fig. 5(B) with R_2 and C_2 removed are shown in Fig. 6(A) for the case where the reactor is short-circuited and in Fig. 6(B) where the reactor is operating. Again the effect of the saturable reactor in distorting the transformer primary voltage is apparent. Without the reactor the primary voltage is, of course, given by equation 6. With the reactor connected an analysis of the primary voltage yields:

$$e_p = E_{p1} \sin (\omega t + \psi_1) + 0.62 E_{p1} \sin (2\omega t + \psi_2) + 0.38 E_{p1} \sin (3\omega t + \psi_3) + 0.27 E_{p1} \sin (4\omega t + \psi_4) + 0.14 E_{p1} \sin (5\omega t + \psi_5) + 0.10 E_{p1} \sin (6\omega t + \psi_6) + 0.10 E_{p1} \sin (7\omega t + \psi_7) + 0.07 E_{p1} \sin (8\omega t + \psi_8) + \dots \tag{11}$$

Similarly, the saturable-reactor waveform of Fig. 6(B) may be written as:

$$e_{sr} = E_{sr1} \sin (\omega t + \theta_1) + 0.65 E_{sr1} \sin (2\omega t + \theta_2) + 0.37 E_{sr1} \sin (3\omega t + \theta_3) + 0.32 E_{sr1} \sin (4\omega t + \theta_4) + 0.15 E_{sr1} \sin (5\omega t + \theta_5) + 0.12 E_{sr_1} \sin (6\omega t + \theta_6) + 0.10 E_{sr1} \sin (7\omega t + \theta_7) + 0.07 E_{sr1} \sin (8\omega t + \theta_8) + \dots \tag{12}$$

The presence of harmonics in the primary voltage of the half-wave circuit or the unsymmetrically loaded biphase circuit requires consideration of the same factors previously discussed for the full-wave case. In addition, another important factor exists which is peculiar to these circuits. For certain d-c saturation levels in the reactor, the instantaneous reactor and line voltages e_{sr} and e_g may combine to produce a peak voltage e_p on the transformer primary which appreciably exceeds that of the line. This condition will exist during the nonconducting half-cycle for the half-wave case and during the lightly loaded half-cycle for the biphase case. The secondary voltage e_s will also exhibit these peaks as will the inverse voltage on the rectifiers which are nonconducting during this half-cycle. In Fig. 5(B), e_2 represents the inverse voltage on these rectifiers when R_2 is large compared with R_1. Comparison of Figs. 6(A) and (B) shows that, with the saturable reactor, the peak primary voltage exceeds the peak line voltage by 30%. The resulting increase in the rectifier inverse voltage e_2 is also shown.

These excessive voltages are likely to

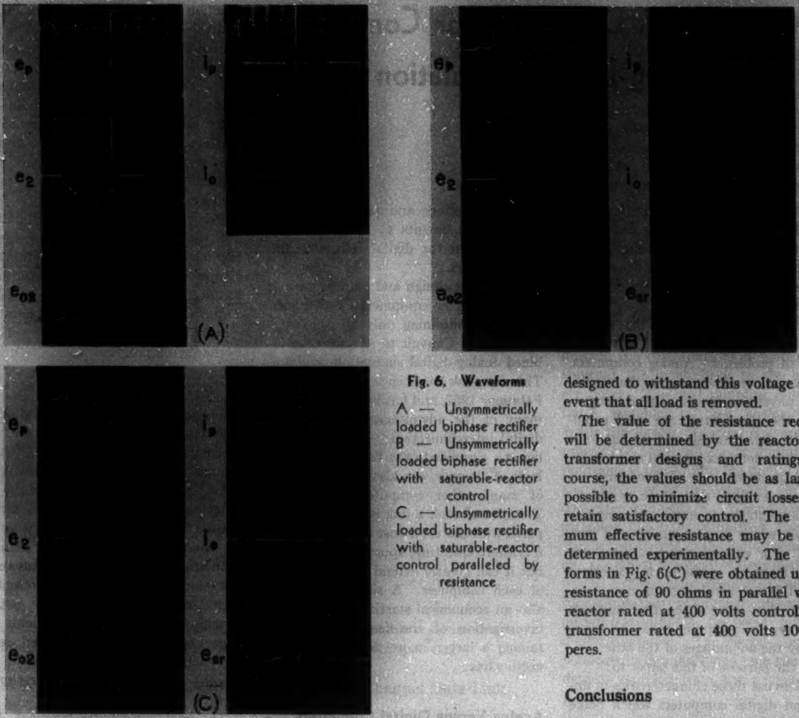


Fig. 6. Waveforms

A — Unsymmetrically loaded biphase rectifier
B — Unsymmetrically loaded biphase rectifier with saturable-reactor control
C — Unsymmetrically loaded biphase rectifier with saturable-reactor control paralleled by resistance

exceed the limits of safe operation for both the rectifier elements and the transformer unless precautionary measures are taken to prevent unbalanced loading of the circuit, unless additional safety factor is provided in the circuit components, or unless means are provided to limit the voltage stresses. A system of interlocked switchgear is one method which can be employed to reduce the probability of operation with one side of the biphase rectifier unloaded. This may not be completely effective, however, since most practical interlock systems can be bypassed and it does not apply to the half-wave case. The second approach to the problem, providing additional safety factor in both the rectifier inverse-voltage rating and the transformer insulation level, can be completely effective. This method is currently being applied successfully to power supplies using vacuum tubes or mechanical rectification where higher inverse-voltage ratings may be obtained with relative ease and reasonable expense. Recently there has been a trend toward the use of the more efficient silicon diodes as rectifier elements. These diodes are appreciably more sensitive to inverse voltage than vacuum tubes and mechanical rectifiers. In addition, larger inverse-voltage ratings cannot be provided as economically. The third method, that of limiting voltage stresses, may be implemented by a minor change in the primary circuit, i.e., by the addition of a resistance in parallel with the a-c winding of the saturable reactor.

The waveforms of Fig. 6(C) were recorded under conditions identical to those of Fig. 6(B) except that the aforementioned resistance was added. The effectiveness of the resistor in reducing the primary voltage peaks may be recognized by comparing the two waveforms of e_p in Figs. 6(B) and (C). Peak primary voltage has been reduced to a value approximately equal to the peak line voltage. As previously discussed in the full-wave case, the rectifiers must be designed to withstand this voltage in the event that all load is removed.

The value of the resistance required will be determined by the reactor and transformer designs and ratings. Of course, the values should be as large as possible to minimize circuit losses and retain satisfactory control. The maximum effective resistance may be easily determined experimentally. The waveforms in Fig. 6(C) were obtained using a resistance of 90 ohms in parallel with a reactor rated at 400 volts controlling a transformer rated at 400 volts 100 amperes.

Conclusions

Care must be exercised in the design of rectifiers using saturable-reactor controls in the primary circuit, particularly if the rectifier circuit and load are such that unsymmetrical currents exist. The resulting excessive voltages can be minimized by paralleling the reactor with a dissipative element. The additional losses introduced are usually small. If they can be tolerated, this method provides a practical alternative to increasing transformer insulation levels and rectifier peak-inverse ratings.

References

1. Introductory Electrical Engineering (book), C. H. Willis, H. M. Chandler. D. Van Nostrand Company, Inc., Princeton, N. J., 1952, pp. 451–56.

2. Magnetic Amplifiers (book), H. F. Storm. John Wiley & Sons, Inc., New York, N. Y., 1955, chap. 4.

3. Frequency Analysis, Modulation and Noise (book), S. Goldman. McGraw-Hill Book Company, Inc., New York, N. Y., 1948, p. 15.

4. Applied Electronics (book), Truman S. Gray. John Wiley & Sons, Inc., second edition, 1956, pp. 316–29.

A 2-Channel Data Link for Combined Analog-Digital Simulation

J. GREENSTEIN
ASSOCIATE MEMBER AIEE

THE NUMBER of computing laboratories using both analog and digital computers is greatly increasing. Each type of computer is normally operated independently of the other, receiving the computing task for which it is better suited. It is generally agreed that neither type of computer is ideal for solving all classes of problems. Analog computers, for example, are widely used for the simulation of physical systems; digital computers, on the other hand, are used extensively for data analysis and for the solution of very complex physical and mathematical problems. It is becoming increasingly evident that there would be many advantages to having a computer available which is part analog and part digital. For example, such a computer would make it possible to simulate sampled-data control systems more realistically. Also some of the disadvantages of one type of computer could be overcome by the advantages of the other.

It is the purpose of this paper to:

1. Discuss those characteristics of analog and digital computers which make it desirable to combine them.
2. Describe the 2-channel link which was constructed at Convair for the purpose of interconnecting its analog and digital computers.
3. Illustrate the function of the data link in a real-time analog-digital simulation.
4. Indicate some of the additional advantages of combining an analog and a digital computer.

The Addaverter System

Convair, in order to interconnect its *550* operational amplifier analog computer with its Sperry-Rand *ERA 1103* digital computer, ordered a special high-speed 25-channel data link from Epsco Inc., Boston, Mass. This data link was installed at Convair in June 1957. It is primarily being used to help simulate the flight performance of the Atlas Intercontinental Ballistic Missile. This data link, called the Addaverter system (a contraction of analog to digital and digital to analog converter system), contains 15 analog to digital converters, 10 digital to analog converters, and gating and channel selection circuits to transfer binary data between the digital computer and the converters.

During the design and manufacture of this Addaverter equipment, a simplified data link containing only two channels was built at Convair to initiate its combined analog-digital simulation program. This data link has been in operation since February 1957, and has by its use proved many of the advantages of an analog-digital interconnection. The author believes a simplified data link, such as was built at Convair, can also meet the needs of many other computer laboratories. The cost of such a link is very low compared with the total cost of the computer it is interconnecting, approximately 0.1%, but it can greatly extend the application of each computer. A simple system is also an economical starting point for the investigation of the feasibility of obtaining a larger, more flexible interconnecting link.

Analog Versus Digital Computers

DIGITAL COMPUTER ADVANTAGES

One of the most important advantages of a digital computer is its ability to compute with essentially unlimited accuracy. On an analog computer it is difficult to limit errors to less than a few per cent and, unfortunately, the accuracy tends to drop as the problem complexity increases. Another important advantage of a digital computer is its high operating reliability. Since a digital computer generally consists of switch-type electronic circuits requiring only a yes-no response, its reliability can be much greater than high-gain analog circuits requiring operation over a wide range of input conditions.

The setup time required just prior to running a digital problem, once the programming has been completed, is extremely short. The complete program is usually conveniently stored on a paper or magnetic tape. All that is required to run the problem is to feed the stored program into the machine. An analog computer problem requires considerably more preparation time. Potentiometers, switches, function generators, etc., must be adjusted and then a complete check must be made for equipment troubles and setup errors. Trouble-shooting of a digital computer is often easier when equipment troubles do occur since the components of the computer and their arrangement remain essentially fixed regardless of the problem. This makes it possible to prepare a separate diagnostic program which can be used for the life of the machine.

Because of the greater dependability of a digital computer, and because less setup time is required for each run, the digital computer is more suitable than the analog computer for production-type problems. These problems involve a large number of repeated runs with essentially no changes in the digital program.

ANALOG COMPUTER ADVANTAGES

Frequently it is desirable to study the characteristics of actual equipment under realistic operating conditions. Often this is done by connecting the equipment to an analog computer and then using the computer to simulate, in real-time, some of these conditions. This fact that an analog computer can be used for real-time simulation is an important advantage. For many real-time problems, present-day digital computers are actually not fast enough. The arithmetic elements of a digital computer are time-shared, requiring each computation to wait its turn thereby introducing undesirable time delays.

The analog computer language, being in the form of a continuous voltage, is more compatible with other analog equipment. If a digital computer is used, costly analog to digital and digital to analog converting equipment is usually required, since communication with a digital computer must be in the form of a binary-coded signal appearing on a number of parallel wires. Also, digital results appear discontinuously, in a point-by-point fashion requiring special extrapolation techniques.

The analog computer's particular suitability for simulating physical systems is another important advantage. The components of an analog computer are easily programmed to operate in a manner analogous to the physical problem. These

---

Paper 60-10, recommended by the AIEE Computing Devices Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., January 31–February 5, 1960. Manuscript submitted June 27, 1957; made available for printing September 24, 1959.

J. GREENSTEIN is with Convair, San Diego, Calif.

The author wishes to acknowledge the assistance of P. Sheretz, W. Vecera, and C. Richards in the design and construction of this data link.

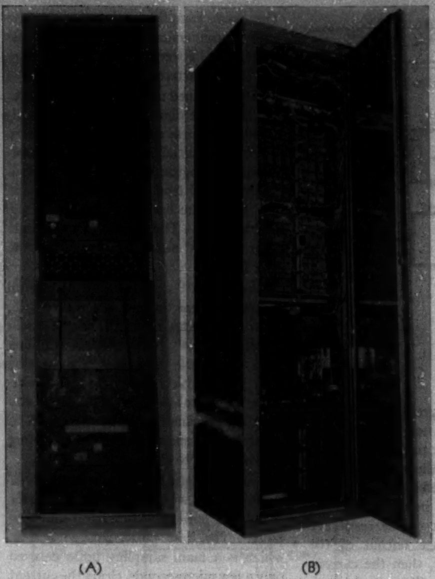


Fig. 1. Convair 2-channel data link

A—Rear view
B—Front view

computers are also often used "open shop," allowing an engineer to solve his own problem directly on the machine. Simple changes in the problem can usually be made by merely switching a few patch-cords. A larger number of steps are required to prepare a physical problem for a digital computer, thus multiplying the possibility of human errors and sometimes introducing frustrating delays. The problems must first be transformed into algebraic terms (possibly by a mathematician); then coded by a programmer to convert the problem into machine language; magnetic or paper tapes are made; finally the problem is fed into the machine by the computer operator. Changes in an existing program require almost as many time-consuming steps. At the present time, some much needed advances are being made in digital programming techniques to enable easier and quicker programming of physical problems.

In summary, it can be stated that the digital computer has an accuracy and reliability advantage and is better suited to perform production-type problems. On the other hand, the analog computer is better able to perform real-time physical simulations involving actual system hardware.

By interconnecting an analog to a digital computer, all of the foregoing advantages are available resulting in a more effective computing system.

Convair Two-Channel Data Link

Function

The simplified Convair data link contains one analog to digital converter and one digital to analog converter; the data link also contains transistor flip-flops, pulse drivers, gates, and delay circuits, to synchronize the flow of binary data between the digital computer and the converters. The complete data link is housed in a standard relay rack as shown in Figs. 1(A) and 1(B). The transistor circuitry, which uses approximately 150 transistors, shown in Fig. 1(B), is mounted in the rear of the relay rack behind the two converters, on hinged panels to enable access to the converters.

Fig. 2 illustrates the function of the link. The input analog voltage from the analog computer is sampled by the analog to digital converter and converted into binary form; the sampling time is controlled by an external sample pulse source. After the conversion is completed, the transistor data transfer circuits transfer the binary data into the digital computer. The digital computer uses this sampled data in its computations, and when it has finished the computations it prepares to transmit the result back to the analog computer. The next sampling pulse initiates sampling of new analog data. While this analog-to-digital conversion is proceeding, the result of the digital computations is placed into the digital to analog converter by the data link and then converted into voltage. When the analog sampling has been completed, this new data is transferred to the digital computer and the whole process just described is repeated. The analog computer performs its computation continuously, although its output is modified by the inputs from the digital computer. The digital computer performs its computations each time new data is sampled from the analog computer.

Digital Computer Input-Output

The digital computer connected to the data link is a Sperry Rand *ERA 1103*, also known as a Univac-Scientific. All data transfer between this computer and the data link is via the computer 36-bit input-output *B* flip-flop register (*IOB*). When "writing" into external devices, such as

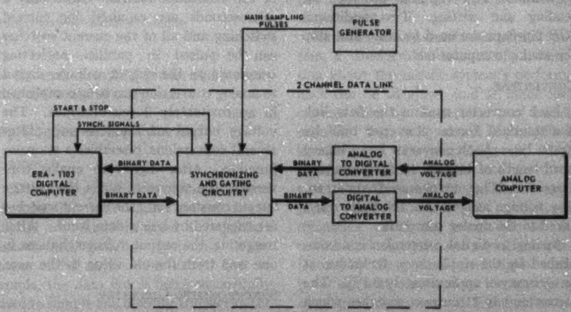


Fig. 2. A 2-channel data link interconnecting an analog and a digital computer

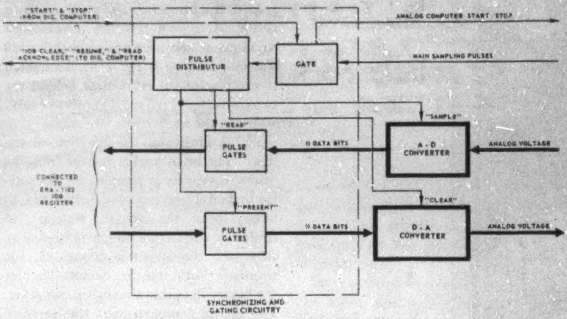

"START" & "STOP" (FROM DIG. COMPUTER)
"IOB CLEAR," "RESUME," & "READ ACKNOWLEDGE" (TO DIG. COMPUTER)
PULSE DISTRIBUTUR
GATE
ANALOG COMPUTER START, STOP
MAIN SAMPLING PULSES
"READ"
"SAMPLE"
PULSE GATES
11 DATA BITS
A - D CONVERTER
ANALOG VOLTAGE
CONNECTED TO ERA - 1102 IOB REGISTER
"PRESENT"
"CLEAR"
PULSE GATES
11 DATA BITS
D - A CONVERTER
ANALOG VOLTAGE
SYNCHRONIZING AND GATING CIRCUITRY

Fig. 3. Synchronizing and gating system

the data link, the digital computer transfers the data word from a storage location in its high-speed electrostatic memory to its input-output register. Then the computer transmits a signal to the external source, indicating that a data word is available in its output register. When the external device has absorbed this data, it must transmit a "resume" signal to notify the computer to perform the next input-output instruction. When the digital computer is programmed to "read" an external device, it waits for the word to be placed into its input-output register. When the external device is ready, the device transmits the binary word to the computer by pulsing the input-output flip-flops and notifies the computer with a "read acknowledge" pulse that the word has been transmitted. The computer then transfers this word from its input-output register to a high-speed memory location and goes on to the next instruction.

The data link has a word length of 11 binary digits; therefore, only 11 of the 36 available *IOB* flip-flops are used for reading and writing. Two additional *IOB* flip-flops are used to start and stop the analog computations.

CONVERTERS

Each converter used in the data link is a standard Datrac converter built by Epsco Inc. Each converter has a binary word length of 11 bits, 10 data bits plus a sign bit. They are operated over an analog voltage range of ± 100 volts referenced to the analog computer.

Analog to digital conversion is accomplished by the ripple-down technique at an accuracy of approximately 0.1%. The converter has 11 current switches which draw current through a resistance matrix. Upon receipt of a sampling pulse from an external source, a programmer turns the current switches on in sequence, starting with the most significant bit (sign bit) and rippling down to the least significant bit (0.1 volt). As each switch is turned on, the analog voltage input is compared to the voltage output of the resistance matrix. If the voltage output of the matrix is found larger, then the current switch is turned off. Since each comparison takes approximately 2 microseconds and there are 11 current switches, complete analog to digital conversion takes approximately 22 microseconds. The binary value of the sampled analog voltage is determined by reading the binary state of the current switches. All the current switches on corresponds to +99.9 volts. All the current switches off corresponds to −99.9 volts.

By a simple modification, each analog to digital converter can be used for digital to analog conversion. Digital to analog conversion is accomplished by turning on the proper current switches. The resulting analog voltage appears at the output of the resistance matrix. Since only 2 microseconds are required for current switching and all of the current switches can be pulsed in parallel, neglecting transients on the output voltage, digital to analog conversion can be accomplished in approximately 2 microseconds. The voltage output of the converter, after several conversions, resembles a staircase function. After digital to analog conversion, the converter holds the voltage output constant until the current switches are triggered by a new data word. After triggering, the output voltage changes in one step from the old value to the new.

SYNCHRONIZING CIRCUITRY

Fig. 3 is a block diagram of the transistor synchronizing and gating circuitry. The main sampling pulses synchronize both the conversion of data and the transfer of data between the digital computer and the converters.

In a combined analog-digital computer operation, the digital computer is first allowed to perform its initial computations. When these initial computations are completed, the digital computer transmits a "start" signal from an *IOB* flip-flop and also places the first data word into the *IOB* register. The computer then waits for a "resume" signal to come from the data link.

The "start" signal from the digital computer opens the gate shown in Fig. 3, allowing the transmission of the main sampling pulses. The first main sampling pulse causes the generation of an analog start signal which places the analog amplifiers into the operate condition, thereby starting the analog computations. The second and succeeding main sampling pulses are fed to the pulse distributor.

The pulse distributor consists mainly of delay circuits which generate a sequence of delayed pulses for each main sampling pulse. The first pulse, labeled CLEAR, which is a main sampling pulse delayed only a few microseconds, clears the current switches of the D-A converter to zero. This causes the voltage output of the converter to start dropping. A "present" pulse follows the CLEAR signal by a few microseconds, and gates the waiting contents of the *IOB* register into the D-A converter current switches, thereby presenting a new voltage to the analog computer. At the same instant that the new data is presented to the analog computer, a sample pulse is fed to the analog to digital converter to initiate analog to digital conversion.

During the conversion process, a *IOB* CLEAR signal from the pulse distributor clears the *IOB* register, readying it for the new sampled data. A "resume" signal is then transmitted to the waiting digital computer, to allow it to continue to the next input-output instruction which will be an "External Read" instruction. The computer then waits for the analog computer's new data word.

When the analog to digital conversion is completed, a "read" pulse from the pulse distributor gates the binary value of the sampled voltage from the A-D converter into the *IOB* register. The data pulses are followed by a "read acknowledge" signal to the digital computer to inform the computer that the data word has been placed into the *IOB* register. The digital computer then continues its computation using the new data word.

At the end of the digital computation,

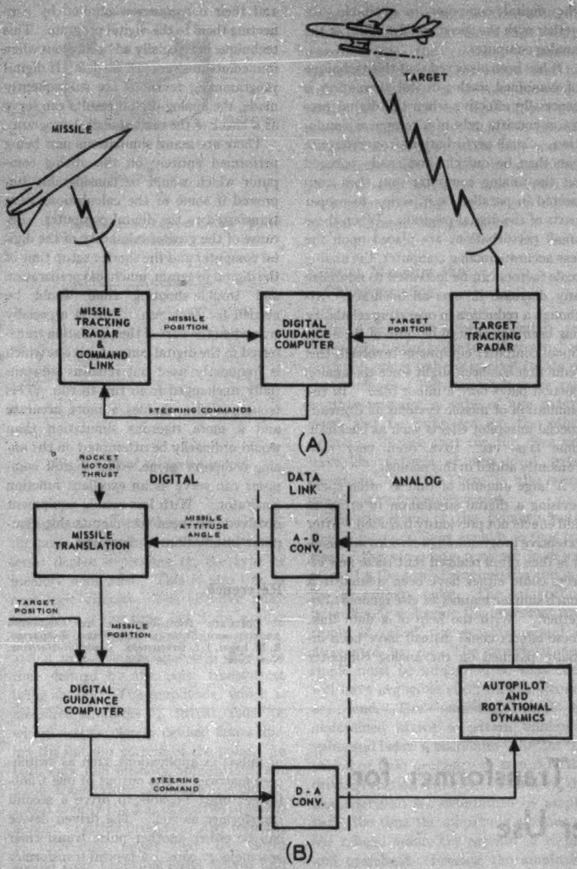


Fig. 4. Simulation of a guided missile system

A—Guided missile system. B—Combined analog/digital simulation

the new result is placed in the *IOB* register and the computer waits for the next main sampling pulse. At the next sampling pulse, new data is exchanged between the computers and the digital computations are again performed. Each digital computation must be completed before the next sampling pulse arrives or the digital computations will fall behind those performed by the analog computer.

This process is continued until the end of the combined analog-digital run. At the end of the run the digital computer places a "stop" signal on a particular output line. This signal causes a halting of the main sampling pulses and a resetting of the analog computer.

Application of the Data Link

General

The primary function of the 2-channel data link is to interconnect the analog and digital computers in order to stimulate mixed analog-digital systems. Many other useful applications exist. For example, the data link is being successfully used to sample and directly transfer to the digital computer, analog data from the analog computer and from magnetic tape units. Sampling speeds of up to 4,000 samples per second have been used. Continuous monitoring of a long digital computation is another useful application of the data link whereby digital computer results are fed to an analog recorder via the digital to analog converter.

The 2-channel data link can be extended easily into a multichannel system by time sharing the converters and the gating circuits. In this case, each separate data output from the digital to analog converter must be placed into an analog hold circuit in order to hold the voltage output constant while the converter is being switched to other channels. An analog sample and hold device may be required for those channels being multiplexed to the analog to digital converter if it is necessary to sample several analog channels at precisely the same instant. After each voltage conversion the binary result must be transferred immediately to the digital computer.

A Real-Time Simulation

Fig. 4(A) shows a somewhat hypothetical sampled-data control system which will serve as an example to illustrate the application of the Convair data link for real-time simulation. The figure shows a guided missile which has been launched for the purpose of intercepting and destroying an enemy aircraft. It is necessary that this missile be guided with extreme accuracy to its manuevering target. Assume that separate radar systems are continuously tracking the positions of the missile and the target. A ground-base digital computer samples the position data from the tracking radars at equal time intervals, and then calculates missile pitch and yaw correction signals. These correction signals are transmitted several times per second to an autopilot located in the missile, causing the missile to maneuver in the direction of the target.

From a realistic simulation of this system, a considerable amount of information about the missile system's performance can be obtained, and at much less expense than by the test firing of actual missiles. Fig. 4(B) shows a simplified simulation of this control system. The simulation can be divided into the three parts: missile translation, guidance computer, and autopilot and rotational dynamics. Both the translation of the missile and the guidance equations performed by the ground-based digital computer are best simulated on the more precise digital computer. A high-gain autopilot, controlling changes in missile attitude, would ordinarily have a fast response characteristic with respect to the over-all guidance loop; therefore, it can best be simulated on the analog computer. Since an actual autopilot might not function any more accurately than the analog computer, the lower accuracy of the analog computer is not necessarily a disadvantage. The separate analog and digital simulations are then joined with the help of the data link. This simulation technique allows the lower-frequency higher-accuracy portions of the system to be performed by the digital computer, while the lesser-accuracy higher-frequency portions are performed by the analog computer.

When the actual autopilot equipment becomes available for test, it may easily be placed into the control loop to replace the simulated autopilot. With the lower frequency effects restricted to the digital computer, a slower digital sampling and computation rate can be used, making it easier for the digital computer to keep up in real-time.

MORE EFFECTIVE COMPUTER USE

In the preceding example, the major function of the data link was to enable a more realistic simulation. Convair's experience with the data link also indicates that a more advantageous utilization of computers results when analog and digital computers are interconnected. For each simulation, the accuracy and reliability of the digital computer is available, together with the speed and flexibility of the analog computer.

It has been observed that this technique of combined analog-digital simulation is especially effective when the digital program consists only of a reference simulation. Small perturbations from reference can then be quickly and easily patched on the analog computer and then connected in parallel, or in series, to proper parts of the digital program. When these small perturbations are placed upon the less accurate analog computer, the analog scale factors can be increased to minimize any decrease in over-all accuracy. Although a reduction in operating reliability has been experienced because of the additional computer equipment involved, this reduction has been slight since the analog portion plays only a minor role. In the simulation of missile systems at Convair, special autopilot effects such as backlash, time lags, etc., have been very conveniently added in this manner

A large amount of time is often spent revising a digital simulation in order to add effects not previously included. After extensive programming revisions are made it is then often realized that these new effects could either have been treated in a much simpler manner or else ignored altogether. With the help of a data link, these effects could instead have been initially patched on the analog computer and their consequences studied by connecting them to the digital program. This technique is especially advantageous when immediate answers are needed. If digital programming revisions are subsequently made, the analog-digital results can serve as a check of the revised digital program.

There are many simulations now being performed entirely on the analog computer which would be considerably improved if some of the calculations were transferred to the digital computer. Because of the greater reliability of the digital computer and the shorter setup time of the digital program, much less preparation and trouble-shooting time would be needed for each run. This is especially true when the part of the simulation transferred to the digital computer is one which is frequently used but remains substantially unchanged from run to run. This technique also enables a more accurate and a more rigorous simulation than would ordinarily be attempted on the analog computer alone. The digital computer can serve as an excellent function generator. With less analog equipment involved, the need for huge analog computer installations is decreased.

Vo
Pt.
19

Reference

1. SIMULATE DIGITALLY, OR BY COMBINING ANALOG AND DIGITAL COMPUTING FACILITIES, R. M. Leger, J. L. Greenstein. *Control Engineering*, New York, N. Y., Sept. 1956.

Specifying a Pulse Transformer for Computer Use

R. R. BLESSING
ASSOCIATE MEMBER AIEE

THE pulse transformer described in this paper is a specialized current-driven device that functions as a precision component in large-scale ferrite-core memories, and also as a switching device in computer logic. Typical transformer characteristics do not satisfy the exacting demands of computer applications. Greater precision in specifying and measuring these components is required along with closer tolerances.

The paper will show how the specifications are arrived at, how the devices are measured, and how these characteristics are related to the application.

How the Device Functions

When used as an input device to large core memories, the transformer must respond to pulses that are trapezoidal in shape. It must faithfully reproduce this waveform at one or more secondary windings. It must also be able to respond with an output to only the first of a train of pulses in applications such as switch-core matrices. The output of one transformer must be able to drive a second transformer as well. The driven device can be either another pulse transformer or a memory core. A typical transformer and an input pulse are shown in Figs. 1 and 2.

Critical Requirements

The above describes the ideal which is difficult to achieve. The most critical requirement of this type of transformer is faithful reproduction of the input pulse at the secondary. In memory applications one is not interested in the output of the memory core when information is entered, but only in the pulse that results when the core is interrogated. The interrogating pulse must faithfully resemble the input to the switch core.

The pulse transformer response must be a true reproduction of the input only during the time the memory core is being switched.

Paper 59-119, recommended by the AIEE Electronics Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 1–6, 1959. Manuscript submitted November 3, 1958; made available for printing November 6, 1959.

R. R. BLESSING is with the International Business Machines Corporation, Poughkeepsie, N. Y.

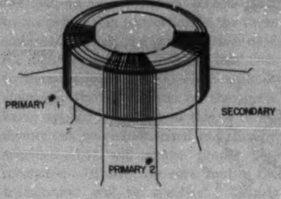


Fig. 1. Specialized current-driven pulse transformer

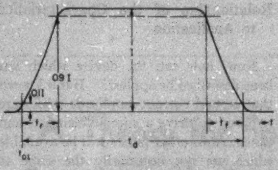


Fig. 2. Input or driving pulse

During the time information is entered, the pulse can be distorted to a greater or lesser degree depending on the type of memory core used. This is also true in core logic circuits. The effective pulse must have a rise time and shape similar to the primary pulse, negligible overshoot, and a negligible droop for a period of time defined by the pulse transformer being driven. The amplitude, which is measured in terms of current, must be within certain closely defined limits during the flat-top portion of the pulse. In both logic and memory applications, the pulse transformer secondary should respond to only the first of a series of pulses applied to either primary. Any response to succeeding pulses that does occur should have negligible effect on the unit being driven. The point of interest is the amount of energy being transmitted to the driven unit expressed in terms of the amount of flux change appearing at the secondary.

Physical Characteristics

Normally, the pulse transformers consist of two or more windings on a magnetic core of toroidal shape. The windings can be placed for maximum or minimum interaction with each other depending on the application. Because most cores are sensitive to shock, it is appropriate to encapsulate them in some plastic compound. The compound selected should be one that does not stress the core in setting. It also must protect the core and its windings from handling damage. The windings are then brought out in a manner which allows their ready insertion into the circuit. In many applications, it has been found best to wire the winding to a base assembly similar to that of the base of a 9-pin miniature tube. By so doing, it can be plugged into an appropriate tube socket. The transformers can also be assembled on a base that will allow ready insertion on a printed wiring card. Fig. 3 shows typical physical configurations.

How to Measure the Characteristics

Fig. 4 shows the actual secondary response of a loaded pulse transformer when the appropriate pulse is applied to one of the primaries. Between each pulse put on that primary, a pulse of like nature must also be impressed on the other primary. Notice that the top does not have the flatness of the ideal pulse shown in Fig. 2. There is also some overshoot and ripple present.

The rise time of the primary pulse should be made short enough so that the rise of the secondary response is determined by the core characteristics and not by the primary pulse. The flat portion immediately following the overshoot and ripple is called the amplitude and is designated by I_m. The overshoot and ripple must be minimized so that they will have negligible effect on the secondary load. This amplitude must be maintained above a certain minimum value and below a maximum value for the period of time necessary to complete the driving function required. Therefore, the characteristics are determined by amplitude, the time the amplitude remains at the critical value, the amount of ripple, and overshoot. Because the amplitude of the input pulse directly affects the switching time, the operating frequency must allow sufficient time for the switching of the core. When the time between pulses approaches the core switching time, the duration of the pulse entering information becomes critical. If the core is not completely switched, it will affect the response during the most critical pulse time.

The pulse response is directly related to the hysteresis loop of the core used. The most important characteristics of the loop are the *Br/Bs* squareness ratio, the high-permeability region, the amount of time it takes a pulse to drive through this high-permeability region, and the rate at which the permeability decreases as saturation is approached. The magnetic core characteristics can be determined by observing the secondary pulse response of the transformer. Fig. 5 shows a typical hysteresis loop. If it is assumed that the core is in the state labeled $-Br$ and a pulse is applied in the direction that tends to switch the core to $+Bs$, the rise time can be adjusted so that it occurs before the loop is traversed to the knee, designated by the numeral 1. The curve from point 1 to point 2 covers a relatively high permeability region. During this time, the shape of the top of the pulse will be governed by the permeability. If this section of the curve is vertical, a flat top response is obtained. It is apparent that the permeability decreases as it approaches saturation at $+Bs$. During this time, the pulse can be expected to droop in direct relationship to the curvature of the hysteresis loop (i.e., decrease in permeability). The period of time during which the amplitude is above a certain predetermined level can deter-

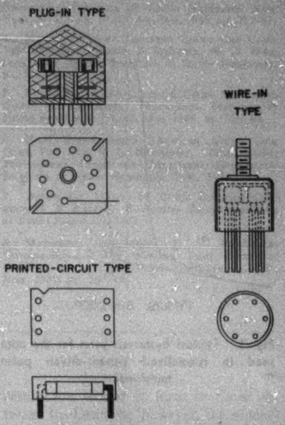


Fig. 3. Typical physical configurations of specialized current-driven pulse transformers

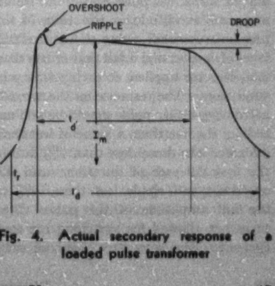


Fig. 4. Actual secondary response of a loaded pulse transformer

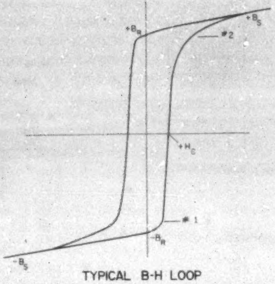


Fig. 5. Typical hysteresis loop for the core used in specialized current-driven pulse transformers

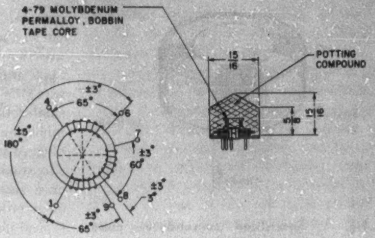


Fig. 6. Typical specialized current-driven pulse transformers

mine whether or not the permeability is satisfactory. The rate of decrease in permeability as saturation is approached can be determined by a measurement of pulse duration at a percentage of I_m.

If a pulse is applied to a primary following a similar pulse applied to the same primary, the portion of the hysteresis loop from $+Br$ to $+Bs$ is traversed. The change in flux that occurs will cause a pulse to appear on the secondary whose amplitude is inversely related to the ratio of Br to Bs. If this ratio is low, the pulse that occurs on the secondary will be large, and vice versa. Since it is desired that no signal appear on the secondary at this pulse time, the response should be small. Therefore, the ratio of residual to saturation flux, which should be kept as large as is possible, can be determined by its relation to the amplitude of this secondary response.

To illustrate further the relationship between the pulse transformer characteristics and the hysteresis loop of the magnetic element, it would be in order to describe a part for a specific application. Fig. 6 shows a drawing of such a unit. Each primary has 17 turns and the secondary 15 turns. A pulse of 400 ma (milliamperes) amplitude, a rise time of 0.2 μsec (microsecond), a duration at the 90% level of 8 μsec, and a fall time of less than 0.6 μsec, are applied at each primary in succession. The response on the secondary during either pulse should be 400 ma ±3%; the rise time, a value of less than 0.5 μsec; the droop less than 5% during the first 2.0 μsec of duration; and the overshoot and ripple less than 10% of the full amplitude of the pulse. Two turns are allowed for switching the core element. The secondary response may vary ±3% of the 400 ma. This allows a total of ±180 ma-turns variation. Therefore, the amount allowed for switching the core is 800 ±180 ma-turns, a variation of coercive force of ±22.3%. The total amount of flux change allowed is controlled by the pulse width at the 50% of full amplitude level. The requirement is that it be 3.5 ±0.4 μsec. Therefore, the core is required to be near the saturation level between 3.1 and 3.9 μsec after the start of the response. Since the total flux change can be achieved with the specified pulses in a definite period of time, the control on the pulse width controls the amount of total flux change.

The high-permeability region is specified by the droop requirement. The pulse should not droop more than 5% within the first 2 μsec after the start of the pulse. During this period of time the core is driven through the high-permeability region of its hysteresis loop and does not start to approach the curvature leading to saturation until after 2 μsec of the pulse have passed.

If only one winding is pulsed successively and the response observed on the secondary, a relatively high voltage response occurs. If this response is integrated, the amount of flux change can be determined which is caused by this pulse traversing the loop from a residual state to its same saturation state. The flux change is related to the amount of energy available for work at the load. By establishing a maximum limit on the integrated response, the Br/Bs squareness ratio is defined. In summary:

1. The amplitude allowed determines the coercive force of the core.
2. The pulse width determines the amount of saturation flux density.
3. The amount of droop determines the permeability of the core in the critical region.
4. The measurement of the flux change from a residual state to the same saturation state determines the squareness of the core.

Relationship of the Characteristics to Application

Now, how can the device which has been described be applied? If it is known that certain current-amplitude pulses are required to drive a specific element, and that certain other pulses will be available which are not necessarily the same as those required for driving this element, a pulse transformer can be devised and specified to meet the application. By requiring a certain pulse amplitude for a specific time the coercive force and the high-permeability region of the core to be used are determined. The pulse width specifies the amount of saturation flux, and the amount of energy the unit being driven can withstand before reacting to it determines the squareness ratio. When the core characteristics are known and are described through the desired characteristics of the pulse transformer, then the special type can readily be specified: the specialized current-driven pulse transformer.

Conclusions

The previous discussion of methods of characterizing specialized current-driven pulse transformers is intended to describe exactly how it is possible to specify the requirements for an application of this component. The method is primarily concerned with the electromagnetic characteristics of the pulse transformer. It has not been pointed out that in conjunction with the characterization that the mechanical and electrical properties are also considered. For instance, insulation resistance, winding resistance, dielectric strength, and the mechanical stability of the unit are determined.

Although the problems involved in characterizing pulse transformers have not been discussed, it might be pointed out that the primary problem is that of achieving the required drive pulse. This is especially difficult if the transformer is a current step-up device. In such cases, the load on the secondary reflects back to the primary in proportion to the turns ratio. Therefore, the current driver used for input would have to stand load fluctuations as great as 500 ohms without changing the amplitude of the pulse being applied. In spite of this, a transformer can be characterized and thereby specified that will assure proper performance. In turn, the core used in the transformer will be specified. This is done without requiring tests of the magnetic characteristics of the core except, indirectly, for its Br to Bs squareness ratio.

It has been found that this procedure is quite satisfactory and that pulse transformers of this nature can be readily described through specification after characterization in this manner.

References

1. Static Magnetic Matrix Memory and Switching Circuits, J. A. Rajchman. *RCA Review*, Princeton, N. J., vol. 13, June 1952.

2. A Magnetic Core Gate and Its Application in a Stepping Register, G. R. Briggs. *Engineering Note E-475*, Digital Computer Laboratory, Massachusetts Institute of Technology, Cambridge, Mass., Oct. 1952.

3. The IBM 705 EDPM Memory System. *Transactions*, Professional Group on Electronic Computers, Institute of Radio Engineers, New York, N. Y., vol. PGEC 5.4, Dec. 1952, pp. 214–23.

4. Magnetic Shift Register Using One Core Per Bit, R. D. Kodis, S. Ruhman, W. D. Woo. *Convention Record*, Institute of Radio Engineers, pt. 7, 1953.

5. Pulse Switching Circuits Using Magnetic Cores, J. Karnaugh. *Proceedings*, Institute of Radio Engineers, vol. 43, May 1955, pp. 570–84.

6. High Speed Shift Register Using One Core Per Bit, V. L. Newhouse. *Transactions*, Professional Group on Electronic Computers, Institute of Radio Engineers, Sept. 1956.

7. Diodeless Magnetic Core Logical Circuits, L. B. Russell. *Convention Record*, Institute of Radio Engineers, 1957.

8. Magnetic Switching, J. A. Rajchman. "Proceedings of the Western Joint Computer Conference," *AIEE Special Publication T-107*, Mar. 1959, pp. 107–18.

A Theory for Space-Charge-Limited Currents with Application to Electrical Precipitation

P. COOPERMAN
NONMEMBER AIEE

THE FLOW of steady electric currents in the interior of any medium is governed by three laws 1. Poisson's law which essentially relates space charge to potential. 2. Conservation of charge. 3. A law of the type of Ohm's law connecting current flow and potential gradient. In most solid and liquid conductors of technical interest, Poisson's law can be satisfactorily approximated by the simpler equation of Laplace, and Ohm's law applies with the conductivity taken as constant. In gases, however, these simplifications cannot be made and hence the problem of calculating such electrical parameters as the current-voltage relationship, field strength, and charge density is much more difficult.

The approximate calculation of these parameters for a corona discharge in coaxial cylinder geometry was done by by Thomson and Townsend.[1,2] Experimental verification of the theoretical formulas has been provided by other investigators.[3,4] However, this geometry is an especially simple one for the reason that all the electrical quantities are functions of the radial distance from the axis and of no other geometrical factor. Where the geometry is essentially 2-dimensional as in electric precipitators of the "duct" type, the methods used in solving the problem for coaxial cylinders do not apply except for rather crude approximations.[5] A very sophisticated attempt to obtain formulas of use in duct-type precipitators was made by Deutsch;[6–7] however, his work was shown to have been in error both by experiment and by theory.[8–10]

The object of this paper is to present a new method for the calculation of the electrical parameters of space-charge-limited currents in gases, and to apply the method to two geometries of interest in electrical precipitation, namely, coaxial cylinder and duct geometry. The results of this work will be good approximations for low current densities and applied voltages not too far removed from the starting voltages. These conditions are, however, typical ones in electrical precipitation. For cylindrical geometry, the results are exactly the classical ones given by Thomson and Townsend. For duct geometry, i.e., a grating of equally spaced wires centered between parallel plates, the results are similar in form to those obtained by Deutsch; however, the numerical coefficients in the formulas are considerably different and, as a result, different conclusions are drawn.

Physical Basis of the Method

A distinguishing feature of electrical conduction in gases is that the electric field which moves the charge also creates it. Thus, when the applied voltage is increased beyond a certain minimum, i.e., the starting voltage, not only does the already existing charge move more rapidly, but also new ions are created at a higher rate. This contrasts strongly with the situation in most solids and liquids where the strength of the applied field acts on ions whose rate of generation is independent of the field. A second characteristic of conduction in gases is that the charge between the electrodes, i.e., the space charge of the moving ions, is comparable in magnitude to the electrostatic charge on the surface of the electrodes. This means that the ions distort the field which both produces and moves them. In solids and liquids, the space charge is usually very small compared with the electrostatic charge and hence the space charge field has a negligible effect on the ions. These features of conduction in gases produce a strong interaction between electric field and space charge which makes it impossible to calculate these two parameters independently of each other. Mathematically, the same difficulty exhibits itself in that the differential equations which govern the phenomena are highly nonlinear.

The basic idea which this paper presents is that the difficulties discussed can be overcome by treating the problem as if no interaction exists but subsequently

---

Paper 60-2, recommended by the AIEE Electronics Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., January 31-February 5, 1960. Manuscript submitted December 22, 1958; made available for printing November 19, 1959.

P. Cooperman is with the University of Pittsburgh, Pittsburgh, Pa.

The author wishes to acknowledge stimulating discussions with H. J. White and J. S. Lagarias.

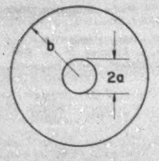


Fig. 1. Cross section of a pipe precipitator

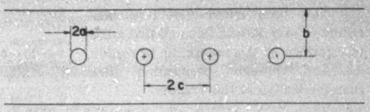


Fig. 2. Cross section of a duct precipitator

making corrections in a suitable fashion. Thus, first it is assumed that there is no space charge but that the entire field is electrostatic in origin, i.e., the field is that of surface charges on the electrodes. With this as a given field, the distribution of the space charge for a given flow of current is calculated. With this distribution of space charge known, calculations can be made to determine how the original electric field has been altered by the presence of space charge. Now this process of changing the field and the space charge alternately is simply repeated. If it could be shown that each repetition brought the results closer to the exact results, one would have a mathematically valid method of solving the appropriate differential equations by an iterative scheme. The proof of this fact is, however, quite complex and hence will be published later in a mathematical journal. In this paper, the reasoning will be on a physical rather than a mathematical basis.

Actually, it will not be necessary to prove the validity of the method in general since only one iteration will be shown here. It will begin with the electrostatic field, then the space-charge distribution corresponding to this field will be calculated, and then the potential corresponding to the space charge recalculated. It is clear that if the curent density is low, the effect of the alteration in the field on the space charge will be small and a second iteration will not be needed. With this restriction, a valid result is obtained. For many situations in electrical precipitation, the limitation is not severe.

Application of the Method

By the mathematical theory given in the Appendix, the potential function $u(r)$ for coaxial cylinder geometry is given by

$$u(r)=\frac{V_0 \log b/r}{\log b/a}+\frac{I \log b/a}{2kV}(b^2-r^2)$$

Notice that the first term is the purely electrostatic potential existing at the starting voltage while the second term gives the change stemming from the space charge. Since at the high-tension electrode $r=a$ (see Fig. 1), the potential must equal the applied voltage V, this equation gives for the current-voltage characteristic

$$I=\frac{2k}{b^2 \log b/a}V(V-V_0)$$

This is the classical formula originally obtained by Townsend. It has been found here again by a much more complicated method, but one which can be extended to the duct precipitator.

Let the dimensions of a duct precipitator be indicated as in Fig. 2. In terms of these parameters, a constant d can be defined as follows:

$$q_1=e^{-2\pi b/c}$$

$$k_1=\left[\frac{2q_1{}^{1/4}\sum_{n=1}^{\infty} q_1{}^{n(n-1)}}{1+2\sum_{n=1}^{\infty} 1^{n^2}}\right]^2$$

$$K_1=\int_0^{\pi/2}(1-k_1{}^2\sin^2\varphi)\,d\varphi$$

$$d=c/k_1{}^{1/2}\,K_1$$

Notice that since k_1 and K_1 are dimensionless, d has the dimensions of length. With this notation, and with the use of the electrostatic potential for a duct given in reference 11, the following is obtained for the modified formula

$$u=-\frac{V_0}{\log(d/a)}\log\left[k_1{}^{1/2}\left|\operatorname{sn}\left(\frac{K_1Z}{c},k_1\right)\right|\right] \tag{1}$$
$$+\frac{\log(d/a)\,I}{kV}(b^2-y^2)$$

where $\operatorname{sn}(K_1Z/c, k_1)$ is a Jacobian elliptic function. From this, the current-voltage characteristic for a duct can be derived as

$$I=\frac{k}{b^2\log(d/a)}V(V-V_0) \tag{2}$$

This formula requires that one be able to compute the constant d, a somewhat formidable task. However, this difficulty can be avoided for it can be shown that the following approximations are valid to within 1% under the conditions indicated

$$\text{for } b/c \leqq 0.6 \qquad d=4b/\pi \tag{3}$$

$$\text{for } b/c \geqq 2.0 \qquad d=c/\pi\, e^{\pi b/2c} \tag{4}$$

For b/c in the range 0.6–2.0, the values of d/c are plotted in Fig. 3. However, if somewhat larger inaccuracies can be tolerated, the value of d given by equation 3 can be used up to $b/c=2.0$. The error caused by this procedure is shown in Table I.

The following results have now been obtained. In a duct precipitator, the current-voltage characteristic is given by equation 2 where the constant d is given by equations 3 and 4 and Fig. 3. The starting voltage, V_0, is given by

$$V_0=aE_0\log(d/a) \tag{5}$$

where E_0 is the "critical" field strength.[12]

Conclusions

Equation 3 states that the length d is independent of the wire-to-wire spacing when b/c is less than 0.6. For a fixed wire-to plate spacing, this means that the current per foot of wire, I, does not depend on the wire-to-wire spacing until the wires approach each other very closely. For example, when b is 4 inches, the current per foot of wire remains constant to within 1% for wire-to-wire spacings up to 13 inches and remains constant up to 15%, for wire-to-wire spacings as close as 4.5 inches. Since there is a normal day-to-day fluctuation in most precipitators of this order of magnitude, it is clear that for practical purposes the current per foot of wire is constant up to the point

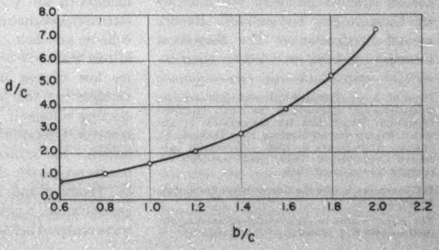


Fig. 3. Calculated values of d/c versus b/c

Table I. Error Made in Calculated Current by Use of Equation 3

b/c	c	log 4b/πa	log d/a	Per-Cent Error
0.8	5.00	4.65	4.70	1.06
1.0	4.00	4.65	4.80	3.12
1.2	3.33	4.65	4.95	6.65
1.4	2.85	4.65	5.10	8.83
1.6	2.50	4.65	5.29	12.1
1.8	2.22	4.65	5.48	15.1
2.0	2.00	4.65	5.69	16.5

where the wire-to-wire spacing approaches the wire-to-plate spacing. This point has been verified experimentally (as described in the author's unpublished company report no. *46*).

Now consider the situation as a and b are held constant while c becomes small (but remains larger than a). Under these circumstances, equation 4 applies. The logarithm in the denominator of equation 4 will usually be small compared with the second term. For example, with $a=0.05$ inch, $b=4.0$ inches, and $c=0.5$ inch,

$$\log(c/\pi a)\cong 1.16$$

$$\pi b/_2c\cong 12.56$$

For these conditions, equation 2 reads

$$I/c=\frac{2k}{\pi b^3}V(V-V_0) \qquad (6)$$

I/c is proportional to the current density at the collecting plate. In fact, if i is the current density at the plate surface, then

$$i=I/4c \qquad (7)$$

and by equation 6,

$$i=\frac{k}{2\pi b^3}V(V-V_0) \qquad (8)$$

This shows that, except for effects due to changes in the starting voltage, close spacing of the wires leads to a constant current density. Beyond a certain point, the total current can no longer be increased by the addition of more wires. In fact, the current will start to decrease as c gets smaller because V_0 will increase. The reason is that for a given configuration, V_0 depends on the field-strength required to initiate a self-sustaining discharge; d changes inversely as c, hence as c decreases, V_0 must increase in order to attain the critical field strength.

Because the current is influenced in opposite directions by the changes in d and V_0 caused by changes in c, it follows that there is a value of c for which the total current is maximum. For $a=0.05$ inch, $b=4.0$ inches, and $E_0=180$ electrostatic units, the current density i is given in Table II. These calculations indicate that the maximum current is obtained for a wire-to-wire spacing of approximately 5.7 inches.

Table II. Current Density as a Function of Wire-to-Wire Spacing

c	i
2.00	172
2.22	197
2.50	208
2.85	213
3.33	204
4.00	194
5.00	170

Note: $k=600$ electrostatic units; $V=150$ electrostatic units.

The wire radius a affects the current voltage ratio primarily through its action on V_0 as shown by equation 5. The effect of the logarithm term in equation 2 is relatively small. Thus, when V is *very* close to V_0, changes in a are very important, but as V increases, these diminish in significance.

For c relatively large, equation 2 indicates that the current varies inversely as $l^2 \log 4\,b/\pi$, whereas for c relatively small, the current is inversely proportional to b^3. Thus, in general, it may be said that if all other dimensions are kept constant, the current varies inversely as b^n where n lies between 2 and 3.

Summary

A new method of deriving theoretical current-voltage relationships and other electrical parameters for single-carrier-current flows in gases has been presented. This method has been applied to corona discharge in precipitators. For coaxial cylinder geometry, the well-known formula of J. S. Townsend was obtained. For duct geometry, essentially new results were obtained. The following conclusions were drawn:

1. For large wire-to-wire spacing, the current per unit length of wire tends to be constant.
2. Except for the effect of the starting voltage, the current density tends to become constant and large for small wire-to-wire spacing.
3. Decreasing the wire-to-wire spacing increases the starting voltage whose effect is to lower the current.
4. Because of the above, there is a value of the wire-to-wire spacing for which the total current is maximum.
5. The wire radius is most important at very low voltages but becomes less significant at higher voltages.
6. The current depends on wire-to-plate spacing inversely as a power lying between 2 and 3.

Appendix

The boundaries of the electric fields to be considered are made up of high-tension electrodes, grounded electrodes, and certain lines of force which are characterized by being also lines of geometric symmetry. These portions of the boundaries will be denoted by B_1, B_2, and B_3, respectively, and the region enclosed by them will be called D. The three laws governing conduction in a gas are

$$\nabla^2 u=-4\pi\rho \qquad (9)$$

$$\nabla\cdot \mathbf{j}=0 \qquad (10)$$

$$\mathbf{j}=-k\rho\,\nabla u \qquad (11)$$

where

$u(x,y)$ = potential function
$\rho(x,y)$ = space charge density
$\mathbf{j}(x,y)$ = vector current density
k = mobility

$\Delta=i\partial/\partial x+j\partial/\partial y$ = vector differential operator with i, j unit vectors in x, y directions respectively. The units are electrostatic.

Equations 1–3 do not have a unique solution unless certain conditions on the boundaries are specified. These are

$$u=V_0 \quad \text{on } B_1 \qquad (12)$$

$$u=0 \quad \text{on } B_2 \qquad (13)$$

$$\partial u/\partial n=0 \quad \text{on } B_3 \qquad (14)$$

where $\partial u/\partial n$ is the derivative of u in the direction of the normal to B_3. In addition to these conditions, the following must be added

$$\rho=f(E) \qquad (15)$$

where E is the magnitude of the field strength i.e.,

$$E=|\nabla u| \qquad (16)$$

The nature of the function f is outside the scope of this theory but can be obtained in other ways. However, in electrical precipitation, it is not necessary to know f since the high-tension electrodes are very small and hence E is practically constant on their surfaces. It follows that equation 7 can be replaced by

$$\rho=\rho_0 \text{ on } B_1 \qquad (17)$$

where ρ_0 is a constant.

The problem above can be broken up into two parts by assumption. Let

$$u=u_e+u_c \qquad (18)$$

where u_e is the electrostatic part of the potential and u_c is the part due to space charge. Then, $u_e(x,y)$ must be the solution of the following equations:

$$\nabla^2 u_e=0 \quad \text{in } D \qquad (19)$$

$$u_e=V \quad \text{on } B_1 \qquad (20)$$

$$u_e=0 \quad \text{on } B_2 \qquad (21)$$

$$\partial u_e/\partial n=0 \quad \text{on } B_3 \qquad (22)$$

Similarly, because of equations 9 and 12–

14, the following conditions must be satisfied for u_c:

$$\Delta_2 u_c = -4\pi\rho \text{ in } D \tag{23}$$

$$u_c = V' \quad \text{on } B_1 \tag{24}$$

$$u_c = 0 \quad \text{on } B_2 \tag{25}$$

$$\partial u_c/\partial n = 0 \quad \text{on } B_3 \tag{26}$$

where V' is some constant. The constants V and V' are to be related to each other by the requirement that for $\rho=0$, $V=V_0$, and $V'=0$ where V_0 is the "starting" voltage of the discharge.

Equations 19–22 are the mathematical form of a perfectly determinate problem, and it is known that there is one, and only one, potential u_e satisfying these conditions. However, equations 23–26 do not determine the potential u_c because the space charge density, ρ, is unknown. The space charge density can be found by combining equations 9–11 with the result

$$\nabla\cdot(\rho\nabla u)=0 \tag{27}$$

At this point, assumption is used and Δu is replaced by Δu_e. With the help of equation 19, this leads to

$$\nabla\rho\cdot\nabla u_e=0 \tag{28}$$

For the 2-dimensional case, $u_e\ (x,y)$ is the real part of an analytic function of a complex variable. Let $v\ (x,y)$ be the imaginary part. Then, it is well-known that

$$\nabla v\cdot\nabla u_e=0 \tag{29}$$

By comparison of equations 28 and 29, it is seen that the general solution for the space charge density is given by

$$\rho=g(v) \tag{30}$$

where g represents an arbitrary function.

The equation of a line of force is $v(x,y)=$ constant, and each line of force corresponds to a different value of the constant. Equation 30 states that ρ is constant along a line of force. Suppose now that ρ is known along a curve which intersects every line of force as, for example, is done by equipotential lines. Then, at any point in the field, one can follow the line of force through the point back to the given line. The value of ρ at the intersection of the line of force and the given line is, by the constancy of ρ, also the value of ρ at the point.

Equation 17 states that ρ is constant along a closed equipotential line. It follows that ρ is constant and the *same* constant on every line of force. By the use of equation 17, therefore,

$$\rho=\rho_0 \text{ in } D \tag{31}$$

Equation 31 makes it possible to find u_c, for equations 23–26 now can be solved without any greater difficulty in principle than offered by the problem of finding u_e. From a practical viewpoint, the usefulness of this method will depend very largely on the ease with which the functions u_e and u_c can be obtained. However, for the two geometries of interest in electrical precipitation, the function $u_e\ (x,y)$ can be found in almost any textbook on the theory of the potential. Furthermore, $u_c\ (x,y)$ can be represented in the form satisfying equation 23

$$u_c(x,y)=\alpha\, u_e(x,y)-2\pi\,\rho_0[Ax^2+(1-A)\,y^2+Bxy+Cx+Dy+E] \tag{32}$$

where α, and A-E are arbitrary constants. It is not generally the case that these constants can be adjusted so as to make $u_c\ (x,y)$ satisfy equations 24–26; however, for cylindrical and duct geometries, this adjustment will be possible, at least approximately.

References

1. CONDUCTION OF ELECTRICITY THROUGH GASES (book), **J. J. Thomson.** Cambridge University Press, Cambridge, England, vol. 2, third edition, p. 552.
2. **J. S. Townsend** in HANDBUCH DER RADIOLOGIE (book), **E. A. Marx.** Leipzig, Germany, vol. 1, 1920, p. 331.
3. LA CHARGE DES PARTICULES SPHERIQUES DANS UN CHAMPS IONISÉ, **M. Panthenier, M. Moreau-Hanot.** *Le Journal de Physique et le Radium*, Paris, France, series 7, vol. 3, 1932, p. 590.
4. A NEW TECHNIQUE FOR THE MEASUREMENT OF CORONA FIELD STRENGTH AND CURRENT DENSITY IN ELECTRICAL PRECIPITATION, **P. Cooperman.** *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 75, Mar. 1956, pp. 64–67.
5. A NEW APPROACH TO THE THEORY OF ELECTROSTATIC PRECIPITATORS FOR USE ON PULVERIZED FUEL FIRED BOILERS, **N. Troost.** *Proceedings*, Institution of Electrical Engineers, London, England, vol. II, no. 101, 1954, p. 369.
6. **W. Deutsch.** *Annalen der Physik*, Leipzig, Germany, vol. 16, 1933, p. 588.
7. ÜBER DIE KORONASTRÖMUNG IN DEN ELECTROFILTERN, **W. Deutsch.** *Physikalische Zeitschrift*, Leipzig, Germany, vol. 34, 1933, p. 488.
8. **N. A. Kaptov.** *Gostekhisdat*, Moscow, USSR, 1947.
9. **V. I. Papkov.** *Electrichestvo*, Moscow, USSR, vol. 1, 1949, p. 33.
10. **L. E. Tsyrlin.** *Zhurnal Tekhnicheskov Fiziki*, Moscow, USSR, vol. 22, 1952, p. 1184.
11. DIE DIFF.-UND-INTEGRALGLEICHUNGEN DER MECHANIK UND PHYSIK (book), **P. Frank, R. von Mises.** Mary S. Rosenberg, New York, N. Y., vol. 2, second edition, 1943, pp. 670–74.
12. DIELECTRIC PHENOMENA IN HIGH VOLTAGE ENGINEERING (book), **F. N. Peek,** Jr. McGraw-Hill Book Company, Inc., New York, N. Y., second edition, 1922.

An Investigation of the Dynamic Switching Properties of 4-Layer Diodes

G. E. McDUFFIE, JR.
NONMEMBER AIEE

W. L. CHADWELL
NONMEMBER AIEE

THE VOLTAGE-CURRENT relation for 4-layer diodes has been described in the literature and is characterized by regions of high and low positive resistance separated by a region of negative resistance.[1,2] The transition between high positive resistance and negative resistance occurs at a potential designated as the breakover or firing potential, V_b. This breakover potential is normally specified in terms of static (d-c) or low-frequency sinusoidal measurements but is, however, a variable dependent on the waveform of the applied voltage. The purpose of this paper is to present experimental data on the variation of breakover potential for various types of applied voltage waveforms, and to give a qualitative explanation of this variation in terms of the breakover process of the 4-layer diode.

Low-Frequency Operation

An idealized voltage-current characteristic of a 4-layer diode is shown in Fig. 1. Although this idealized characteristic does not show the details of the transition from the low-current region to the high-current region, it is closely approximated by the actual device when the characteristic is measured by d-c or low-frequency sinusoidal technique. The slope of the curve of Fig. 1 in the low-current region is of the order of 10^9 ohms and, in the high-current region, is of the order of 10 ohms. The static breakover potential V_{bdc} can be made to vary from 10 to 100 volts by proper physical design. The diodes to be described in this paper have static breakover potentials between 25 and 40 volts.

Moll, Tanenbaum, Goldey, and Holonyak[3] have described the current characteristic of Fig. 1 by the expression

---

Paper 60-19, recommended by the AIEE Semiconductor Metallic Rectifiers Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., January 31–February 5, 1960. Manuscript submitted February 24, 1959; made available for printing November 23, 1959.

G. E. McDUFFIE, JR., and W. L. CHADWELL are with the Catholic University of America, Washington, D. C.

The authors wish to acknowledge the encouragement of Dr. William Shockley and the assistance of Joseph Popolo.

A portion of the material included in this paper is based on a thesis submitted by W. L. Chadwell to the Catholic University of America in partial fulfillment of the requirements for the Master of Electrical Engineering degree.

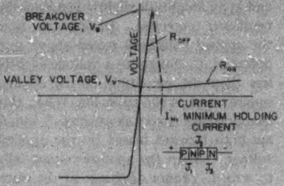


Fig. 1. Static characteristics

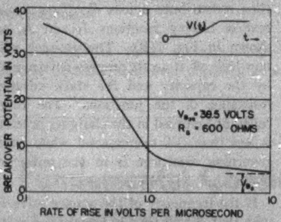


Fig. 2. Variation of breakover potential with the rate of rise of applied voltage

$$I = \frac{I_{co}}{1-(M_p\alpha_1 + M_n\alpha_2)} \tag{1}$$

where M_p and M_n are the hole and electron multiplication factors caused by an avalanche process at the center junction J_2, α_1 and α_2 are the current gain factors of the p-n-p and n-p-n sections of the 4-layer diode, and I_{co} is the current which would flow if there were no avalanche or transistor action. McKay[4] has shown that the avalanche multiplication factors in a single silicon junction diode are approximately equal and a function of the junction potential. Assuming that M_p and M_n are equal, the multiplication factors of equation 1 can be expressed by

$$M_{p,n} = \frac{1}{1-(V_{J2}/V_a)^\beta} \tag{2}$$

where V_{J2} is the potential across J_2, V_a is the avalanche breakover potential of J_2, and β is a constant dependent on V_a. It is also well known that the current gain factor of a silicon transistor increases with increasing current.[5-6] Thus increasing the applied voltage across the 4-layer diode increases both the M's and α's and at the device breakover potential V_b, the quantity $M_p\alpha_1 + M_n\alpha_2$ of equation 1 approaches unity. Under static conditions the current gain factors in the low-current region are small except in the neighborhood of V_b, and the device breakover potential is approximately equal to V_a. If it is assumed that V_{J2} is equal to the applied voltage, equations 1 and 2 may be solved for the breakover potential V_b. However, in considering the dynamic behavior of V_b for various applied waveforms it is more convenient to examine the variations of the M's and α's separately.

Single Transient Breakover

If the breakover potential of the 4-layer diode is measured using waveforms other than low-frequency sinusoids or slowly varying d-c potentials, a wide variation in the breakover potential is observed. Fig. 2 shows the variations of V_b when a variable rise-time ramp function is applied to the diode. Experimentally the ramp function was obtained by using a variable rise-time square-wave voltage source operating at a repetition rate of approximately 1 cps (cycle per second). Because of the low repetition rate and the fact that the switching of the 4-layer diode occurs in a time equal to or less than the shortest rise-time shown in Fig. 2, the voltage waveform initiating breakover can be assumed to be a ramp function. At low rates of rise V_b is approximately equal to the $V_{b_{dc}}$ shown on the static characteristic curve. As the rate of rise increases the breakover potential decreases until a minimum breakover potential V_{b_0} is reached.

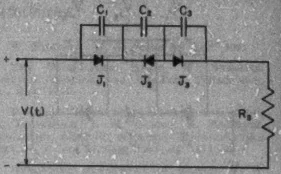


Fig. 3. Descriptive equivalent circuit

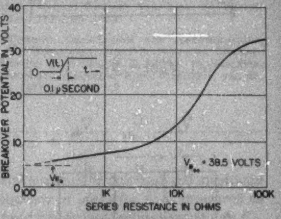


Fig. 4. Variation of breakover potential with series resistance

The behavior of V_b with rate of rise of the ramp function is best understood in terms of the descriptive circuit shown in

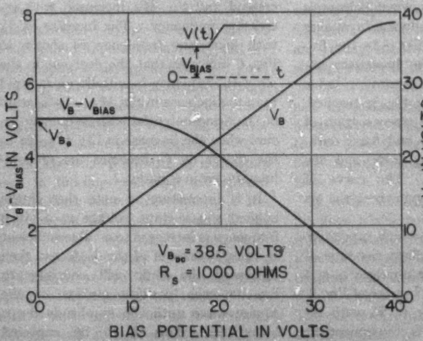


Fig. 5. Variation of breakover potential with d-c bias

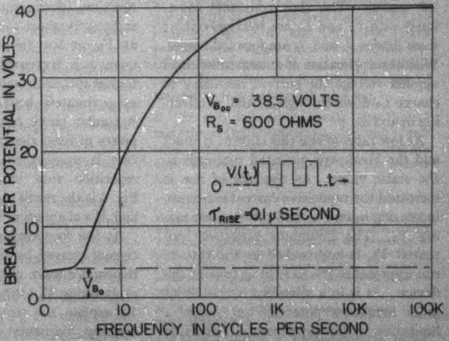
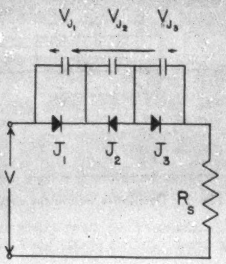
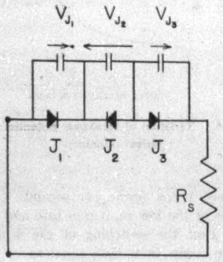


Fig. 6. Variation of breakover potential with frequency of applied square wave



(A) V(t) = V



(B) V(t) = O

Fig. 7. Descriptive equivalent circuit for an applied square-wave voltage

Fig. 3. The capacitances shown in this figure represent the normal depletion layer capacitances of the junction and are of the order of 100 μμf (micromicrofarads) each. R_s is the total series resistance and includes the internal impedance of the source. The time constants associated with C_1 and C_3 are relatively short since diodes J_1 and J_3 are forward biased. With the application of a ramp function of applied voltage the current necessary to charge C_2 flows through J_1 and J_3 affecting α_1 and α_2.

At low rates of rise this current is small and the breakover potential conforms to the static value. As the rate of rise is increased the capacitive current increases, increasing α_1 and α_2 and thereby decreasing V_b. The minimum breakover potential V_{b_0} is approached as the rate of rise approaches the switching time of the device. It is in a sense not a true minimum breakover potential but rather a limitation of the experimental technique employed. If the rate of rise is increased past the maximum values shown in Fig. 2, the breakover potential will continue to fall and eventually the device will resemble a simple capacitor.

Fig. 4 shows the variation of V_b with series resistance; this curve was measured using a ramp function of fixed slope corresponding to the rate of rise producing the minimum breakover of Fig. 2 and at a 2-cps repetition rate. The increase shown in V_b with increasing R_s is also explained in terms of the capacitive current varying α_1 and α_2 as in the preceding argument, with R_s now controlling the magnitude of the capacitive current.

D-c Bias Breakover

When a d-c bias is applied to the diode, and the breakover potential measured using a ramp function of applied voltage with a fixed rate of rise, the breakover potential varies as shown in Fig. 5. The rate of rise of the ramp used for this plot again corresponds to that causing the minimum V_{b_0} in Fig. 2. Since the d-c bias has been applied to the diode at a rate sufficiently slow to preclude breakover by the bias, V_b increases with increasing bias voltage. The curve of $V_b - V_{bias}$ in Fig. 5 shows the effect of the bias potential on M_1 and M_2. As the bias is increased the M's increase as indicated in equation 2, and the increase in α_1 and α_2 by capacitive current necessary to cause breakover is decreased.

Square-Wave Breakover

The applied voltages of the preceding sections have been essentially single transient voltages because of the low repetition rate used. When a square-wave voltage of fixed rise-time is applied to the diode a variation of V_b with frequency results, as shown in Fig. 6. The rise-time of the square-wave voltage is sufficiently short to produce a breakover at V_{b_0} at low frequencies. As the frequency is increased the breakover potential increases until a saturation value approximately equal to $V_{b_{dc}}$ is reached. A similar curve results when a train of pulses of fixed rise-time and fixed duration is applied to the diode and the repetition rate varied. The curve of Fig. 6 is the result of varying both the α's and M's of equation 1.

At low frequencies the high capacitive current caused by the short rise-time of the square-wave voltage increases α_1 and α_2, lowering V_b to the V^*_0 value of Fig. 2. To explain the increase in V_b with increasing frequency it is convenient to examine as a function of time the voltages across the junctions at a fixed frequency and at an amplitude slightly less than the breakover potential. During the time the applied voltage is raised above zero potential, C_2 charges to a voltage slightly less than the amplitude of the applied square wave as shown in Fig. 7(A); this charging time is very short since J_1 and J_3 are both forward biased by the applied voltage. When the applied voltage falls to zero, C_2 cannot discharge at the same rate as the charging rate since J_1 and J_3 are reverse biased in respect to the discharge current. The charge on C_2 must then redistribute across C_1, C_2, and C_3, reverse biasing junctions J_1 and J_3, as shown in Fig. 7(B). The charges will now leak off at a rate roughly determined by the capacity and the static reverse resistance of the junctions. The time constant involved in the charging process is less than 0.1 microsecond while that governing discharge is in the order of several hundred milliseconds.

If the frequency of the square wave is such that the off-time of the applied voltage is less than the discharge time of the junction capacities both the M's and the α's are affected. The charge on C_2 will raise the values of the M's in a fashion similar to that of the d-c bias previously discussed. The charge on C_1 and C_3 will lower the values of the α's in that the initial capacitive current occurring when the square wave rises from zero to the applied amplitude will no longer be a current injected into the base regions of the p-n-p and n-p-n sections of the diode but rather a current which discharges C_1 and C_3. Thus the initial capacitive current does not tend to raise the α's as is the case when C_1, C_2, and C_3 are completely discharged.

Increasing the frequency allows less time for the junction capacitances to discharge; thus the α's are effectively decreased and the M's increased with increasing frequency. The increase in V_b with increasing frequency as shown in Fig. 6 indicates that the decrease in the α's with frequency is initially more rapid than the increase in the M's. Saturation of the firing potential with frequency occurs when the increase in the M's due to the charge on C_2 becomes the predominant factor in equation 1.

It is interesting to note that if the applied square-wave voltage at a given frequency is increased to a value sufficient to cause breakover of the diode and then decreased, the diode will continue to fire on each positive portion of the square wave until the amplitude drops below V_{b_0}. This is to be expected since, when the diode conducts, the voltage across it drops to a very low value;

the major part of the applied voltage is dropped across the series resistance, and C_2 no longer charges to the amplitude of the applied voltage. Each positive going portion of the square wave is now independent of the previous positive portion and the firing potential behaves as if the frequency had been reduced to a very low value. The dashed curve of Fig. 6 shows the variation of V_b with frequency for the case of a conducting diode.

Conclusions

The results of these measurements indicate that the breakover potential of a 4-layer diode is dynamic in nature and dependent upon the applied waveform. Under dynamic conditions, the capacitive current plays a considerable role in determining the potential at which transition between low and high current occurs, and under certain conditions the voltage existing across the center junction prior to the application of the terminal voltage affects the breakover potential.

The curves which have been presented in this paper are typical of the commercial, low-power 4-layer diodes tested although the sensitivity of V_b to the various parameters discussed can be varied to some extent in the design of the device.

References

1. W. Shockley. *U. S. Patent no. 2,655,609.*

2. THE FOUR-LAYER DIODE, W. Shockley. *Electronic Industries and Tele-Tech,* Philadelphia, Pa., Aug. 1957.

3. P-N-P-N TRANSISTOR SWITCHES, J. L. Moll, M. Tanenbaum, J. M. Goldey, N. Holonyak. *Proceedings,* Institute of Radio Engineers, New York, N. Y., vol. 44, Sept. 1956, pp. 1174–82.

4. AVALANCHE BREAKDOWN IN SILICON, K. G. McKay. *Physical Review,* New York, N. Y., vol. 94, May 1954, pp. 877–84.

5. DIFFUSED EMITTER AND BASE SILICON TRANSISTORS, M. Tanenbaum, D. E. Thomas. *Bell System Technical Journal,* New York, N. Y., vol. 35, Jan. 1956, pp. 1–22.

6. CARRIER GENERATION AND RECOMBINATION IN P-N JUNCTIONS AND P-N JUNCTION CHARACTERISTICS, C. T. Sah, R. N. Noyce, W. Shockley. *Proceedings,* Institute of Radio Engineers, vol. 45, Sept. 1957, pp. 1228–43.

L. 79
1
60

Fundamental Considerations of Power Dissipation Limits in Some Bistable Transistor Pulse Circuits

H. RAILLARD
NONMEMBER AIEE

THE PROBLEM of determining the minimum power that has to be dissipated by a transistor flip-flop covering a set of given specifications is one of general interest. Among the various significant considerations which determine the power level are: output power requirements, stability, temperature variations, and required speed of operation. The following considerations deal mainly with the problems of speed limitations at various power dissipation levels. The investigated circuits employ silicon devices to insure good temperature stability and are based upon a design which offers good stability and a constant output voltage. The required increase in power level which is necessary for a desired increase in output voltage is also discussed.

Two essential properties of transistors used in flip-flop operation are current amplification and cut-off frequency. If a single figure of merit is wanted, the product of the two, viz., the gain-bandwidth product, may be used. For applications at low power levels, both current amplification and cut-off frequency have to hold up at low emitter currents and at low collector voltages. The collector saturation resistance indicates how well the gain holds up at low collector voltages and must be as low as possible.

The design procedures discussed here deal exclusively with nonsaturating circuits in order to eliminate the transient delay variations caused by saturation. This is desirable because a wide range of pulse repetition frequencies will be used, including frequencies where saturation delay is significant and where it is not. This paper deals with the nonsaturating emitter-coupled and base-return configurations shown in Figs. 1 and 2, which are widely used. A constant collector voltage swing of 2 volts is assumed as a practical minimum value. For reliable and nonsaturating operation the minimum collector supply voltage E_{cc} is found to be 6 volts for emitter-coupled flip-flops and 4.5 volts for base-return flip-flops.

In order to investigate a number of flip-flops whose power dissipation varies from less than 1 mw (milliwatts) to more than 100 mw, flip-flops of both the emitter coupled and base-return types are designed with various load current swings. The power dissipation P_d is found to be approximately 15 mw per milliampere load current swing.

Each flip-flop is triggered through a suitable gating circuit and the maximum frequency of operation f_m is measured as a function of the coupling capacitance C_K. The results are plotted and shown in Fig. 3(A) for the emitter-coupled circuits and in Fig. 3(B) for base-return flip-flops, with the power dissipation P_d of each flip-flop as a parameter.

The curves shown in Figs. 3(A) and (B) permit the following conclusions:

1. For a given power level the maximum frequency of operation f_m peaks at a frequency f_{mm}. At this frequency C_K is approximately $C_{K\min}$, the minimum operational coupling capacitance, below which the flip-flop does not operate.

2. For a given power level and for $C_K \gg C_{K\min}$ the maximum frequency f_m is inversely proportional to C_K. In this region of C_K values, f_m is determined by the discharge time constant of C_K only.

3. For a given fixed capacitance $C_K \gg C_{K\min}$ the maximum frequency of operation f_m is approximately proportional to the power level of the flip-flop.

4. The locus of the peak maximum frequencies for each power level form the dotted lines of Figs. 3(A) and (B). For larger power levels (larger than 80 mw), f_{mm} becomes independent of increases in circuit power dissipation and is only a function of transistor gain-bandwidth, $\omega_{\alpha B}\,\alpha_B$.

5. At very low power levels (less than 5 mw) the peak frequencies f_{mm} for each power level occur at the same value of C_K and are approximately proportional to the power level of the flip-flop.

Paper 60-20, recommended by the AIEE Electronics Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., January 31–February 5, 1960. Manuscript submitted August 24, 1959; made available for printing November 20, 1959.

H. RAILLARD is with the General Electric Company, Syracuse, N. Y.

The author wishes to acknowledge the helpful discussions with Mr. J. J. Suran, Mr. B. K. Eriksen, and Mr. E. A. Fisch, of the General Electric Electronics Laboratory.

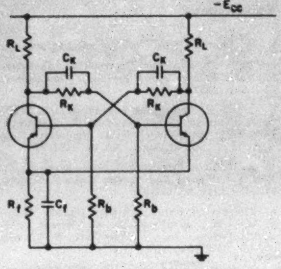


Fig. 1. **Emitter-coupled flip-flop**

An analytical approach to determine the maximum operating frequency f_m of the flip-flop is made as follows.

The time constant T_f of the exponential collector response is

$$T_f=(1+\omega_{\alpha E}\alpha_E R_L C_c)/\omega_{\alpha E} \qquad (1)$$

where

α_E = current amplification (common-emitter)
$\omega_{\alpha E}=\alpha_E$—cutoff frequency
R_L = collector load impedance
C_c = collector capacitance

Also the coupling network time constant T_K can be computed from

$$T_K=C_K R_K R_b/(R_K+R_b) \qquad (2)$$

Equations 1 and 2 are consolidated into a tentative formula for maximum theoretical frequency of operation

$$f_m=1/(T_f+R_K) \qquad (3)$$

The frequency f_m in equation 3 is calculated for all the flip-flops and plotted next to the measured curves of Figs. 3(A) and (B). The correspondence between experiment and theory is in general very good.

From a comparison of the calculations of maximum frequencies f_m described before with the experimental curves f_m in Figs. 3(A) and (B), it is found that the curves peak consistently at a point where

$$T_K/T_f \cong 0.4 \qquad (4)$$

This point, experimentally, is always close to the point where regeneration is lost completely. When the quotient (T_K/T_f) becomes smaller, too little of the collector pulse is transmitted through the coupling capacitor C_K and regeneration is lost. Equation 4 indicates the minimum tolerable value of (T_K/T_f) for regeneration.

The dotted curves in Figs. 3(A) and (B) show a breakover between a region where the peak maximum frequency f_{mm} is approximately proportional to power level P_d, at low power levels, and a region, where f_{mm} is a constant, independent of power level.

By taking the set of equations 1, 3, and 4, and eliminating both T_K and T_f a theoretical value for the peak maximum requency f_{mm} is obtained

$$f_{mm}=4.4 f_{\alpha E}/(1+\omega_{\alpha E}\alpha_E C_c R_L) \qquad (5)$$

in which the denominator consists of two terms, namely, the constant 1 and a term which is inversely proportional to power level. At very low power levels (less than 5 mw) the second term in the denominator is of dominant significance, therefore f_{mm} is proportional to power level in this region. At large power levels (more than 80 mw) the first denominator term dominates so that f_{mm} is independent of power level in this region.

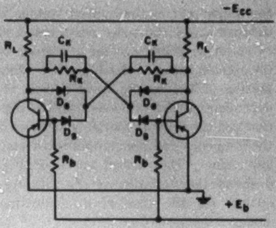


Fig. 2. **Base-return flip-flop**

Breakover between the two regions occurs when the two terms are equal, that is,

$$\omega_{\alpha E}\alpha_E R_L C_c=1 \qquad (6)$$

which happens at a power level of approximately 15 mw.

The measurements and calculations discussed show that in general the following circuit parameters are of influence on the maximum frequency of operation f_m:

1. The current amplification α_E of the transistor.
2. Parasitic transistor parameters, that is collector capacitance C_c in conjunction with circuit load impedance R_L.
3. Cross-coupling elements: R_K, C_K.
4. The gain-bandwidth product of the transistor $\omega_{E\alpha}\,\alpha_E$.

The peak maximum frequency f_{mm} for a certain power level P_d is determined by the following parameters, whereby the numbers presented here are valid for nonsaturating flip-flops of the emitter- and base-return type, with a fixed voltage swing of 2 volts using silicon transistors.

1. $P_d<0.5$ mw. In this region the current

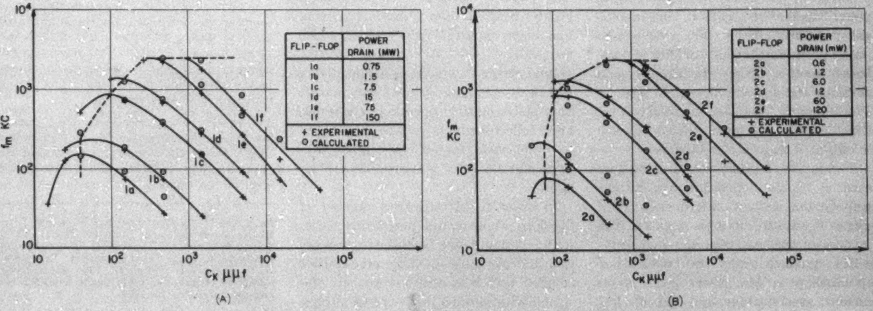


Fig. 3. **Maximum operating frequency f_m versus coupling capacitance C_k for**

A—Emitter-coupled flip-flops 1a–f. $E_{cc}=6$ volts, $\Delta V=2$ volts
B—Base-return flip-flops 2a–f. $E_{cc}=4.5$ volts, $\Delta V_e=2$ volts, $E_b=0$

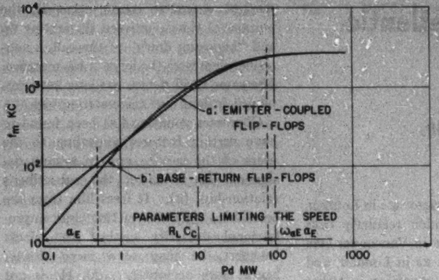


Fig. 4. Peak maximum operating frequency f_m versus power dissipation P_d for emitter-coupled and base-return flip-flops, $\Delta V_c = 2$ volts

amplification α_E decreases rapidly with decreasing power level and limits the frequency range.

2. $0.5\text{ mw} < P_d < 80\text{ mw}$. In this region the frequency range is limited by the collector capacitance-load impedance product R_L C_c.

3. $P_d > 80$ mw. For large power levels the frequency range becomes independent of all stray and circuit parameters and depends only on the gain-bandwidth product ($\omega_{E\alpha E}$) of the transistor.

The peak maximum frequency of operation f_{mm} is plotted in Fig. 4. These curves indicate the various parameters that limit the frequency of the flip-flops.

In a practical case one may want to increase the collector voltage swing to some value larger than 2 volts. For the same load current swing, the supply voltage, the resistance value of the circuit, and the total power dissipation increase proportionally to the voltage swing.

The very low power region discussed before, where the flip-flop speed is limited by the decreasing current amplification α_E, is determined by the emitter current, thus the power dissipation limit of this region will increase proportionally with voltage swing.

The large power region where the flip-flop speed limit becomes independent of circuit parameters is determined by the load impedance level. Therefore the value of power dissipation limiting this region will increase proportionally to the square of the voltage swing.

References

1. Design of Non-saturating Flip-Flops, J. J. Suran. *Internal Report*, General Electric Company, Schenectady, N. Y., Jan. 10, 1957.

2. Transistor Circuit Engineering (book), R. F. Shea, et al. John Wiley & Sons, Inc., New York, N. Y., 1957.

3. The Effect of Collector Capacity on the Transient Response of Junction Transistors, J. W. Easly. *Transactions*, Professional Group on Electron Devices, Institute of Radio Engineers, New York, N. Y., vol. PGED-4, 1957.

4. Two-Terminal Analysis and Synthesis of Junction Transistor Multivibrators, J. J. Suran. *Ibid.*, Professional Group on Circuit Theory, Institute of Radio Engineers, vol. PGCT-3, 1956, p. 26.

A Character-Metered Transatlantic Switching System

I. S. COGGESHALL
FELLOW AIEE

P. HOLCOMB, JR.
MEMBER AIEE

IN ADDITION to offering public telegram services over transatlantic facilities, cable and radio carriers today offer various forms of temporary, switched teleprinter interconnections between correspondents, upon request of a calling party, in a manner and through means characteristic of long-distance telephone calls. Regardless of forms of the national teleprinter exchanges involved, the international portion of the haul is usually between manual boards at central offices. The placement of a specific call over a particular idle two-way trunk among those at the moment available is controlled by concerted action of the two operators guided by information exchanged over a parallel order circuit or "bid channel" connecting them.

These operators reach their respective calling and called subscribers through tie lines terminating in the jacks of manual switchboards or "turrets." Connection of the tie lines with an overseas trunk is accomplished by inserting, into the jacks, plugs led to the trunks by "cords" into the circuits of which all manner of signaling, ringing, and supervisory controls may be built. These controls may contain as many automatic features as desired, such as for testing the tie line for continuity and "line-busy," or for clocking and recording the call. Owing to differences between the engineering problems presented by the typed and spoken word, an extraordinary amount of automation is included in the cord-plug-jack circuits of the telegraph system described.

System Aspects

INTERNATIONAL METERED COMMUNICATIONS

The IMCO (International Metered Communications) system departs operationally from telephone practice in the following particulars which specify the designs:

1. Any one patron in New York (for example) does not require random interconnection with all patrons in London (and vice versa); rather, subscribers pair themselves off for contract service. Hence, when MA in New York signals for service it is almost invariably for connection to joint contractor MA in London. Similarly, MB in London regularly calls MB in New York; RF in Vancouver (via New York) deals with RF in London, and so on. The typical connection in either direction may therefore be completely described by 2-letter calls, like MA, MB, RF; a circumstance of which full advantage is taken in design.

There is an exception, however. MA New York can be connected to MB London through an operation called "cross bidding." For the present, cross bidding may be ignored.

2. The relative absence of cross bidding results in the constant availability of far-end tie lines to callers. If MA New York is not sending to MA London, probably no one is. The important circumstance that there are virtually no tie-line busy conditions encountered on the IMCO system substantially improves the speed of service. Means are incorporated in the system to handle counter bids in the event that MA New York and MA London happen to call each other at the same instant.

3. All initiated calls are one-way (not back-and-forth) calls. A reply to a query (if a reply in the opposite direction is called for) or an acknowledgment of receipt is separately initiated as a call upon termination of the query. Transatlantic messages on IMCO, it has been found, like those sent over duplexed private-wire networks in the United States, are one-way communications rather than the question-and-answer type. By virtue of this fact, not only is the trunk load factor improved by eliminating entirely the idle half of a duplexed setup for every call, but also it is made possible to utilize the benefits of reperforator storage at central office transmitting points without the usual mutual interference encountered from opposite ends of a half-duplexed reperforator facility. Keeping trunks fully loaded is, of course, of economic benefit to all concerned.

4. In contrast with widespread teleprinter switching practices in Europe and the United States of charging by the minute or of measuring service by the nickel, both of which are based on clocked time, IMCO charges are based upon straight character count; that is, the number of times, between the setting up and "knocking down" of the call, a subscriber depresses the keys of his machine. The same tariff device has long been employed in land-line telemeter service.

Character counting has been found to have certain features appealing to the users of the service and to system designers. In terms of the subscriber's relationship: (*a*). It frees him from the economic whip of pitting his unprofessional dexterity against the clock. Inexpert, he may send more slowly, hence more accurately. (*b*). He is not required to pay for "units or a fraction thereof," but only for the service used. He may send messages as short and as often as he wishes at no greater cost than if he accumulates them to send in batches with less frequent calls. Coded messages of half a dozen characters or less are not unusual. The letter R followed by a terminating signal suffices to acknowledge receipt of a message.

In terms of telegraph system design, character counting contributes to the designer's freedom to utilize the trunk efficiency inherent in a reperforator storage system; and since it is unnecessary to have a calling party hold the line until a through end-to-end circuit can be established, he is enabled to afford each calling customer an instantaneous start or "go-ahead."

5. It is necessary to cater to the relative incapabilities of people who attempt to operate customers' teleprinters. The best of them are experts, but even they, after special training and practice, must exercise care in transmitting (along with the text of messages, which is comparatively simple) the information which identifies a message in its sequence, in time and date, and in other respects covered by the preamble, or "heading," of a telegram. In the IMCO apparatus as much as possible of such information is generated electromechanically at the central office and transmitted to the

---

Paper 58-1143, recommended by the AIEE Telegraph Systems Committee, and approved by the AIEE Technical Operations Department for presentation at the AIEE-IRE Second National Symposium on Communication Systems, St. Petersburg Beach, Fla., December 3–5, 1958. Manuscript submitted September 29, 1958; made available for printing June 25, 1959.

I. S. COGGESHALL is with AIEE Headquarters and P. HOLCOMB, JR., is with The Western Union Telegraph Company, New York, N. Y.

The authors thank their associates, A. W. Dickey, H. F. Krantz, G. T. Fontaine, E. A. Wold, R. I. Mills, and J. H. Browne, for assistance in the development of the equipment and circuits and in the preparation of the paper. The apparatus was engineered and manufactured by the Taller & Cooper Division of American Electronics from functional designs and specifications prepared by Western Union.

customer right on his tie line as the lead-off of, and in a manner of speaking, an advance receipt for, his message. The advantages of having an automatic machine intervene in the preparation of this difficult and necessary but purely routine information are obvious: it can be prepared more rapidly and accurately than any introduced manually, and it can be made constantly and instantaneously available.

6. Complete automatization was to be avoided because of the expense of duplicating automatically the discretionary functions of the assigned operators, particularly in the switching of inbound trunks to tie lines. However, in the design of the turrets and the cord-transfer system, every effort was made to make the operators' jobs easy, straightforward, and devoid of fatigue. It is believed that it has been made impossible for them, without being "defeated" by interlocks, to insert a right or wrong plug in a wrong or right jack. Automation that is included sounds alarms and blocks faulty manipulation and switching failures of trunks, tie lines, and equipment. The accuracy of all plugged-in connections is checked by machine before trunk transmission is permitted to begin. Customers' calls appear twice in succession to guard against the cardinal fault of sending a message into the wrong tie line.

7. For special reasons having to do with cost and time limitations, only standard communication apparatus was to be used. There are therefore no units of equipment requiring description in this paper.

Operational Environment

Following is a brief résumé of operational routines and physical equipment layouts.

PLACING AND CONFIRMATION OF BIDS

New York subscriber MA wishes to place a call to his London correspondent MA. His printer is not yet running. He pushes a button which energizes a signal light over the MA jack in the New York cable-office turret. The turret is shown in Fig. 1.

The IMCO operator selects an idle sending cord, say R, and plugs in.

At once, the calling customer gets his cue to begin sending; his printer motor starts, and within 6 seconds the heading of his message is automatically typed on his printer: RFH526MAMA. This sequence uniquely identifies the message he is about to send from all others which will pass through the IMCO system. He and his London correspondent may use it for reference and service purposes.

Fig. 1. IMCO turret, jack field, and cords

With the heading now on his blank, the customer may begin typing his message at his leisure. The heading and text are simultaneously being stored temporarily at New York cable office in a standard *FRXD* type of reperforator transmitter.

At the end of the text the patron depresses the FIG and J keys (bell signal) to terminate the call. His part is finished.

Meanwhile, at the central office, the mechanism which generated the message heading for the customer also passed it over the transatlantic bid channel for the London IMCO operator to act upon, but in the abbreviated form RMA. The London operator, seeing this notation on his inbound bid printer, inserts his receiving cord R in the MA jack as directed.

This accomplishes several things at once: 1. It tests the London MA tie line for continuity and starts the subscriber's printer. 2. It establishes a through message facility from the New York reperforator to the London customer's printer. 3. It sends the combination RMA back to New York, where an automatic comparator checks it for accuracy against the RMA sent from there. If anything in the mean time has gone wrong (wrong cord, wrong jack, call garbled going through to London or returning) the bid will be rejected and have to be remade from New York. 4. It makes a heading appear on the London customer's receiving blank, reading slightly different from that now in storage in New York, viz: RFH53-MAMA.

Meanwhile, normal operation of the comparator at New York will have released the heading and text of the message temporarily stored in the reperforator there and it will begin to run over the R transatlantic message trunk (not the bid trunk) into the London subscriber's printer. The original New York heading will appear right below the London-generated heading, thus:

RFH53-MAMA
RFH526MAMA

and the subscriber, once having been initiated, can trace from these notations the genesis of the call and by accurate record its transit time. The second notation is still the one used for reference.

Following the text, when the New York sender's terminating bell signal reaches the New York and London operators in succession, the customers' printer motors will be stopped, the operators pull down the cords, and the call is complete.

As soon as the customer's printer has stopped, he may initiate a new call by pressing his calling button even though the operators have not yet unplugged the call just completed.

TIE-LINE TERMINATIONS

The subscribers' teleprinters are standard type *15s* in consoles, with a-c motors started and stopped by relays controlled from the central offices. The tie lines themselves are operated on a single-line-to-ground basis. Tie-line power comes from the central office. Termination there is in the IMCO jack turret, which has been made an adjunct to a standard plan *2* concentrator already

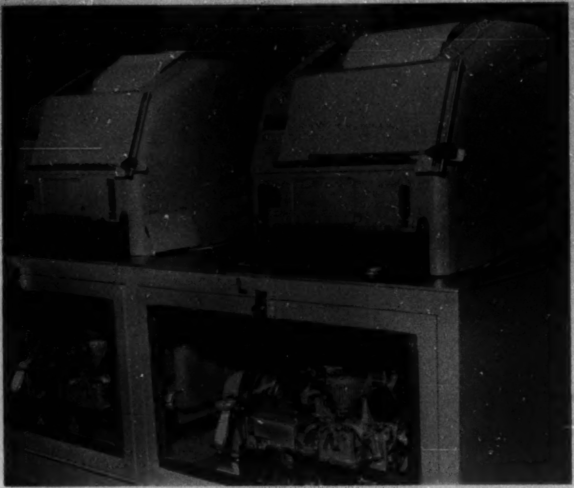


Fig. 2. Monitor cabinet, FRXDs, printers, and tape reels

described in the literature.[1] The concentrator contains signal circuit relays responsive to a push button in the subscriber's office, a busy-test feature, test for open line, and control of the printer motor, as stated. In teleprinter practice, as distinguished from telephony, the starting, stopping, and testing of the subscriber's printer is important to the dependability of the service.

A westbound call cannot be "sent to the wind" in New York if the called party's tie line or printer relay stands open or busy because of safeguards built into the plan *2* tie-line concentrator used.

A further safeguard has been tried out and is under consideration for automatically pretesting an unattended receiving printer for full operational capabilities at the instant the tie-line connection is made. Due to time-zone differences, unattended operation is frequently encountered in transatlantic practice. This test would consist in detecting, at the moment of plugging in at the central office, the customer's printer's response to a specially transmitted bell signal, FIGS J. The FIGS character would be transmitted with a spacing bias, the J with a marking bias, both within the range tolerances of a properly adjusted printer. If the line and equipment added bias to one side or the other, the tolerance for safe operation would be exceeded and the test would be negative. If the test proved positive, the bell lever system in the printer would trip a line-opening contact. This responsive momentary open-line condition immediately after transmission of the FIGS J combination is the positive test sought and detected at the central office to denote a properly operative printer.

If the printer did not respond satisfactorily to this test, because its motor was not running at proper speed, range, and synchronism, a bid could not be confirmed and the message could not be sent. This test has been found experimentally to be good irrespective of the length of the tie line or the number of repeaters in it. When applied to IMCO, it is expected to be more stringent and quicker than the test given printers by means of the "who are you" key-and-cam test used on Telex machines in Europe. The revelation of printer identity is, of course, made superfluous by the precautions taken in IMCO to check the identity of the tie line in which the sole printer operates.

CENTRAL OFFICE LAYOUT

At the central cable office the IMCO turret with its jack field of 200-tie-line capacity, its shelf equipped with a double row of 20 sending and 20 receiving cords, and the position occupied by the IMCO operator with his bid and monitor printers, dominate the installation. Surrounding this group as its physical core are 20 monitorial cabinets in which are mounted 20 inbound monitor printers and 20 outbound *FRXD* reperforating monitor units, connected with the respective cords (see Fig. 2). Automatic tape winders make it possible to refer to any transmission in either direction.

In an inconspicuous position is mounted the IMCO control rack, the heart and brains of the installation, containing the character generators, the program switch, time and date group generators, the comparator unit, an *FRXD* collector unit, relays, and miscellaneous equipment, all included in Fig. 3.

Character Generating and Programming

In order that an intelligible sequence of Baudot character signals may be presented as needed to the tie lines and reperforators as message headings, to the bid trunks and printers as bids, and so on, a multipoint stepper-type rotary[2] machine switch is employed which is called the program switch. This switch is stepped by a character generator[3,4] in such a way that the program switch stops on each of its step points just long enough to pick up in rapid succession the seven units of the Baudot start-stop teleprinter code presented on that point during that particular sweep of the switch.

The presentation of an appropriate generated character to each successive switch point at any given sweep of the switch is the function of other machine switch and relay circuits, among which are the cord-designating contacts, the date-and-time generating assembly, the message-sequence designator, and the call-letter designator.

These controllers of generated characters are shown in the lower part of Fig. 3 and will be described. Meanwhile, their relation to the typical message heading already referred to: RFH526MAMA, may profitably be considered. R designates the cord plugged up by the IMCO operator; F designates the day of the week (SMTWHFA); H is the hour (*A* to *Z; I* and *O* omitted), and *52* the minutes, of the time of day. The figure *6* is the sequence designator, the system message serial number associated with the designated time of day. MA is the designation of the calling and called parties; it is repeated for safety as MAMA.

It will be obvious to those familiar with printer functions that any such succession of printed characters has to be supported by accompanying printer "stunt" functions. In this case they are: LTRS CR LF RFH FIGS 526 LTRS MAMA CR LF.

To meet all needs, the entire alphabet of 32 characters is continually produced

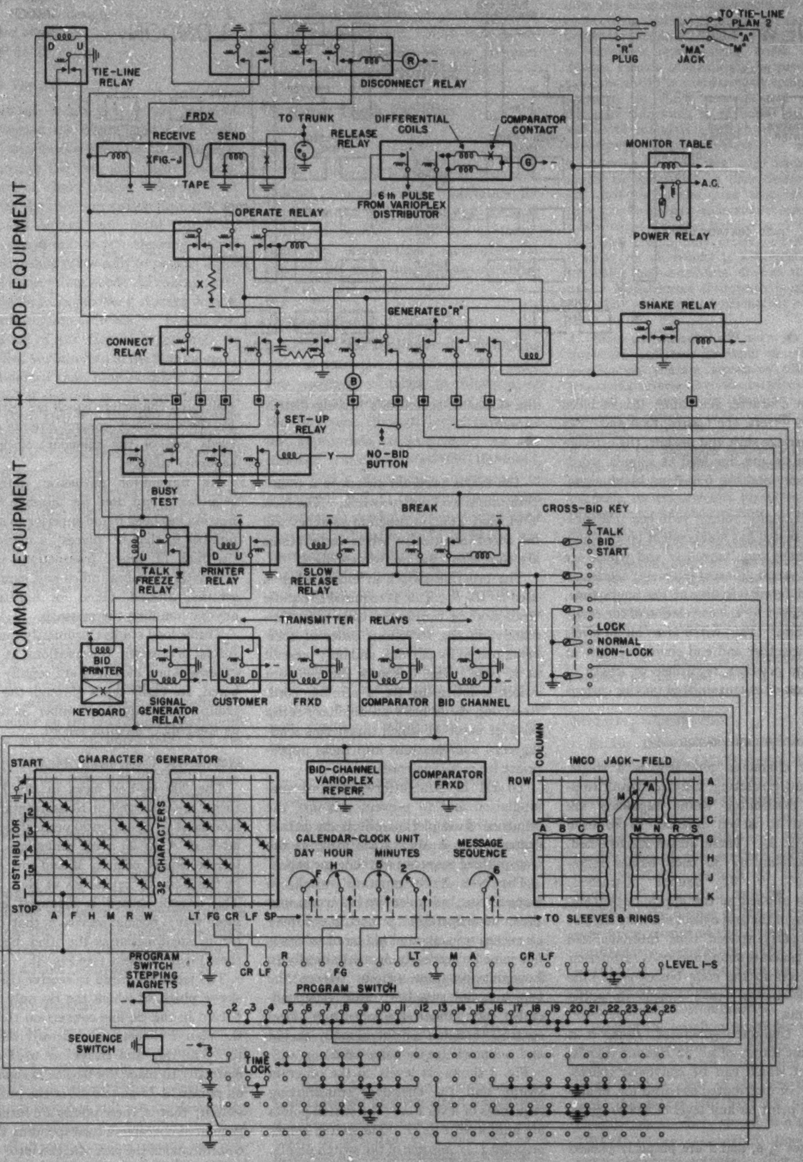

TO TIE-LINE PLAN 2
"R" PLUG
"MA" JACK
"A"
"M"
TIE-LINE RELAY
D
U
DISCONNECT RELAY
R
FRDX
RECEIVE
SEND
FIG.-J
TAPE
TO TRUNK
RELEASE RELAY
DIFFERENTIAL COILS
COMPARATOR CONTACT
G
6th PULSE FROM VARIOPLEX DISTRIBUTOR
MONITOR TABLE
A.C.
POWER RELAY
OPERATE RELAY
X
GENERATED "R"
SHAKE RELAY
CONNECT RELAY
B
CORD EQUIPMENT
COMMON EQUIPMENT
SET-UP RELAY
Y
NO-BID BUTTON
BUSY TEST
BREAK
CROSS-BID KEY
TALK
BID
START
LOCK
NORMAL
NON-LOCK
TALK FREEZE RELAY
PRINTER RELAY
SLOW RELEASE RELAY
TRANSMITTER RELAYS
BID PRINTER
KEYBOARD
SIGNAL GENERATOR RELAY
CUSTOMER
FRXD
COMPARATOR
BID CHANNEL
BID-CHANNEL VARIOPLEX REPERF.
COMPARATOR FRXD
COLUMN
ROW
IMCO JACK-FIELD
A
B
C
D
M
N
R
S
G
H
J
K
START
CHARACTER GENERATOR
DISTRIBUTOR
1
2
3
4
5
STOP
32 CHARACTERS
A F H M R W
LT FG CR LF SP
CALENDAR-CLOCK UNIT
DAY HOUR MINUTES
F
H
5
2
MESSAGE SEQUENCE
6
TO SLEEVES & RINGS
PROGRAM SWITCH STEPPING MAGNETS
CR LF
R
FG
LT
M A
CR LF
LEVEL 1-5
PROGRAM SWITCH
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25
SEQUENCE SWITCH
TIME LOCK

Fig. 3. Schematic diagram, sending-cord circuit

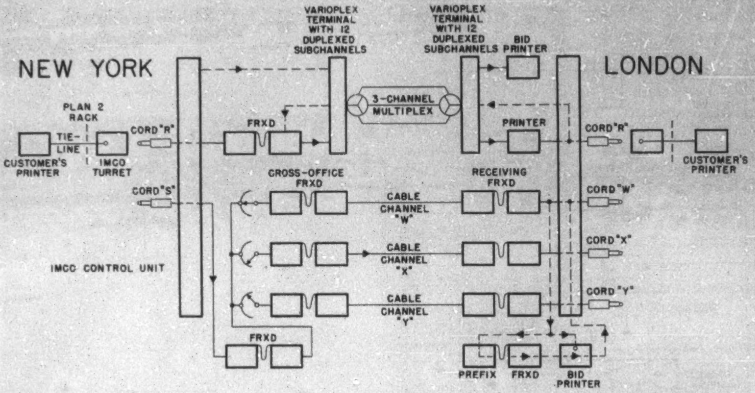


Fig. 4. Schematic diagram, IMCO transatlantic facilities

by the character generator; this includes the 26 letters, the figures, FIGS and LTRS for carriage shift and unshift, the carriage return CR and line feed LF signals, word-space SP, and the 5-spacing blank signal BK.

Like many rotary switches, the program switch has independent levels which are swept simultaneously and which are used to send different character sequences into different circuits at the same time, as required by the correlation of the entire operation. The switch arms on all levels start together and end simultaneously at a home position, regardless of length of the sequences transmitted on the various levels.

Character Generator

The character generator is a single, constantly rotating, standard 7-unit stop-start distributor whose segments are applied to seven busbars in rotation (see Fig. 3). These busses are overlaid, grid-fashion, by 32 code busses representing the 32 Baudot characters. Diodes applied at the appropriate intersections of the grid collect marking battery at properly spaced time intervals and hence generate simultaneously 32 different characters in the code busses each time the distributor brush sweeps its seven segments.

The character generator, then, is a constant source of all 32 Baudot signals, each of which can be conveyed continuously, or distributed as called for, to any switch point on any level of the program switch.

Pulses *3*, *4*, and *5* are properly phased with respect to the generated characters to actuate the program switch, advancing it one step per completed revolution of the character generator's driving distributor.

Program Switch

The single program switch is a multipoint multilevel rotary switch. The New York and London program switches will be wired identically when installations are completed and are so described.

The program switch at either office is used by the operator as required to handle eastbound or westbound calls indiscriminately, in the fortuitous order of their being placed by patrons, and only one call at a time. Fig. 3 shows the program switch connections for a sending cord. Receiving cords have an auxiliary setup relay at point *Y*, which transforms sending cord combinations into those appropriate for receiving cords.

To aid in straightforward cord manipulation and to prevent mix-ups, the sequence of completing calls at the distant central office is always the same as the transmitting sequence at the home end.

The New York program switch, in transmitting bids over the bid trunk into the London operator's printer, does so over an exclusive eastbound bid trunk which is not joined electrically or mechanically to London's program switch. Hence the two program switches rotate independently, each under the operators' local control, though used functionally to bid and acknowledge the same calls.

Fig. 3 shows six levels on the program switch and the character transmitting sequences of level *1-s*. The switch points are numbered *1* to *25*. The switch homes on point *1*. On some of the switch points, characters picked up from the character generator are never changed by secondary switching. Other points pick up generated characters which are permuted by the secondary control circuits of the cords and of the calendar, clock, and sequence units.[5]

An unpaid-for authentic date-time-sequence group used for message identification has been found superior to a serial number in that it "places" a message in the flow of timed transactions in the originating business office and cannot be errored, altered, faked, or omitted by anyone handling the message.

Traffic load studies, ordinarily made by special live-traffic observations at extra expense when messages are serially numbered, are easily made through the time-group and sequence-number notations on the IMCO monitor record.

Automatic Call Identification

The 200 tie-line jacks in the IMCO turret are arranged in 10 horizontal rows labelled *A* to *K*, and in 20 vertical columns labelled *A* to *Z*, with omissions. The intersection of column *M* and row *A* will be the locus of tie-line MA at New York. The London board is arranged in the same way; normal service is that called for, in one direction or the other, between these identical jacks (see Fig. 3).

The jacks are wired to receive tip-ring-sleeve plugs, of which the tip only is involved in the tie-line connection through the plan *2* concentrator already referred to. The ring and sleeve are used automatically to place the tie-line designation on points *14-15* and *16-17* of the program switch; that is, these points are furnished generated characters corresponding to the co-ordinates of the jack plugged into. The sleeve of the plug picks up the letter locked to that particular column of sleeves, and the ring picks up the letter locked to that particular row of rings. Together, the adjacent switch points signal the 2-letter tie-line designator. The call is repeated, as an accuracy check, on points *16-17*.

SHAKE RELAY

As to any single jack or any group of jacks within the turret jack field, additional information may be made accessible through the ring and sleeve grid by the use of shake relays. During such times as the program switch does not require call letters, multipoint 2-position shake relays may divert the ring sleeves from revealing office calls to passing other bits of information about the calling party obtainable from auxiliary storage devices and subprogram switches.[6,7] In this way the name of the city of origin may be identified and inserted in the transmission; the standard time zone at origin may be included if desired, or Eastern Standard Time translated into the time in the zone of origin; priorities may be inserted in messages where authorized, as in the case of government tie lines; or serial-numbering[8] machines may be introduced, or special tests may be applied to particular tie lines. In Fig. 3 a shake relay is shown as a means of simplifying circuitry, avoiding a multiplicity of vacuum tubes or the overloading of diodes.

COMPARATOR FRXD

Level *4* is used to set up in a comparator *FRXD* the combination: LTRS RMA, to await the bid confirmation.

The arrival of London's switching apostrophe on the bid channel westbound causes contacts on upper-case S of the bid printer to close, starting stepping pulses to the comparator *FRXD* transmitter. The previously stored LTRS RMA combination from level *4* will be stepped in unison. If the two sequences test for identity, the eastbound message reperforator, containing by now the original heading and part of the text, will be unlocked and will begin transmitting on the overseas trunk. If the identity test fails, the operator is called in on alarm to initiate a rebid (compare reference 9).

DOVETAIL BIDDING

Since the character generator spins at 390 rpm, it puts the program switch through its 25 steps in only 4 seconds. Nevertheless, when bidding is heavy, the program switches will not be developing their capabilities for handling calls at full effectiveness unless means are provided for an operator to set up a second call during the 4-second interval the program switch is preoccupied with the first. This is done by rendering a cord associated with a second call, inactive, even though plugged up, until the program switch is on its home position. With such overlapping, full capacity of the program switch may be realized during peaks.

Overseas Trunk Facilities

BID CHANNEL

The bid channel is a duplexed varioplex channel, equipped with teleprinters for each direction and at each end, kept at priority disposition of bids and confirmations but automatically available for supervisory wires and service conversations at other times.

TYPES OF CABLE FACILITIES

Although the sending cords, upon being plugged into jacks, automatically cause the program switches to function as described, their main function is to join the tie lines, through reperforation, to an appropriate transatlantic one-way message trunk (see Fig. 4). As fitted, New York's 20 sending (eastbound or outbound) and 20 receiving (westbound or inbound) cords, associated with their opposites in London, will join 40 one-way transoceanic facilities for independent communication to subscribers.

A switch on each cord permits its equipment to be moved from a varioplex to a multiplex channel or vice versa. Both types of facility employ 5-unit transmission on the main trunk lines of cable which operate near the ceiling of capacity. Both also are joined to 7-unit start-stop teleprinters on the tie lines.

UTILITY OF VARIOPLEX FOR IMCO SERVICE

The steady capacity of both the multiplex channels and varioplex subchannels under normal total system loading is in excess of the output of the average customer's operator using an on-line start-stop keyboard. Manipulation of varioplex and multiplex capacity preserves this desirable relationship.

However, the following advantages of the varioplex[10-12] over the multiplex make it the preferred facility for IMCO service (the multiplex handles the overflow and fall-back):

1. Ten times as many customers can simultaneously be given direct connections.
2. Whenever a customer's call is acknowledged by allocation of a cord and he is given the automatic heading as a signal to go ahead, transmission on the message trunk begins from the reperforator immediately and results in the two subscribers working directly together. Only the varioplex utilization of full trunk capacity permits the economical assignment of direct connections indiscriminately to slow or deliberate senders. Under single-channel or multiplex reperforation, trunk transmission cannot economically be permitted to begin until the customer has completely prepared his entire message and signed off, thus subjecting it to initial delay as well as to cumulative delays behind other customers' unitized messages.
3. Messages originating over each particular tie line arrive at their destinations via varioplex in the same sequence that they start, whereas a succession of reperforator-switched messages handled over several multiplex trunks may arrive out of turn if delay on all trunks is not equal or if expensive oldest-call-first devices are not used. Maintenance of sequence is essential in dealing with buy-and-sell market transactions.
4. Since the varioplex will not work unless duplexed, and since its chain circuits and subchannels contain means of efficiently employing reverse-direction interlocks, it is adapted to automatic checking of the accuracy of switching in the manner already described.
5. Since the varioplex incorporates automatic machine switching of multiple subchannels among fewer available multiplex channels, its use promotes maximum trunk utilization for switched services. This is an advantage where facilities are limited and other services have to be maintained.
6. The pattern stop built into the varioplex, tests for and recognizes line trouble and abruptly stops transmission in both directions, constituting a main-line transmission error-detecting device which multiplex does not possess.
7. Because a varioplex subchannel preoccupies no line time while it is inactive, lightly loaded bid channels are feasible.

MULTIPLEX OPERATION

In the cable plant, multiplexes predominate, so the IMCO design contemplates their auxiliary use to greatest advantage. If no more varioplex trunk sending cords are available, the IMCO operator uses a "multiplex" switch to modify one of the sending cords, say cord S. Unlike cord R in varioplex, heretofore considered, modified cord S conditions the release of the stored message upon receipt of the knockdown bell marking its completion. The message is then handled cross-office as a unit. An IMCO heading is furnished the sending tie line automatically.

Each of the overflow multiplex sending channels contains an *FRXD* and a collector switch with seeker and sender levels. When clear of the last message sent, the seeker will look for a "live" message, cross-office, stored in a monitor *FRXD* whose associated knockdown relay has been actuated. The sending level will complete the circuit between the two reperforators cross-office, and the message will start flowing over the outbound trunk.

At the receiving office the switching sequences dictated by fortuitously arriving unit messages over the several channels are funneled into one page printer for operating convenience. Successive messages are identified by the usual sending headings. To facilitate switching them at the IMCO turret, the receiving IMCO operator has a prefix-storing *FRXD*, emptying prefixes (only) into his inbound bid printer (see Fig. 4).

Imagine the system to be operating at London. When an inbound overflow cord is unplugged from its circuit-completing "home" jack, or from a receiving IMCO customer's tie line in response to a disconnect signal, the prefix-storing *FRXD* is thrown into temporary association with the bid printer, permitting the prefix (only) of the next message to be displayed there for switching purposes (but without sacrificing it). The cord which the operator has just unplugged is thereupon placed in the new jack so indicated. The second message then proceeds into the tie line. The operator thus plugs up new connections for the entire multiplex channel group into the tie-line turret at the continuing instruction of his single-bid printer, without having to identify the channels he is switching. This is believed to be a novel way safely to switch to tie lines a volume of messages pouring in over a trunk group.

Character Counting and Billing

A number of years' experience has been had with character counting and automatic meter reading, on-line and off-line. Either method provides the carrier with an accurate character count automatically and immediately upon knockdown; if on-line, the information may, if desired, be communicated into the tie lines.

A point in favor of off-line counting, that is, saving reperforated tape and running it later through high-speed local printers, is that copy so prepared can be tailor-made to the bookkeeping and billing operation. For example, short uniformly-sized tapes, consisting of preamble and a fixed amount of reperforated text in the form of a toll ticket, can thus be prepared for sorting by customers' accounts. Bills may be actually prepared on page printers, automatic totalizing being also performed in one monthly operation. On-line copy is not so readily adaptable.

As fitted to IMCO at the time of writing, an off-line printer is connected to a simple counting program switch and a pair of cross-checking 4-stage decade counters. In the home position of the switch the counters are restored to zero. The switch steps once on the FIGS combination of the prefix and a second time on LF. It then releases the character counters but continues to step until a fixed quota of LFS has been received. When FIGS J goes through the transmitter, the switch supplies any extra LFS that may be needed to form a blank of manageable size, then prints the independent but coinciding readings of the two counters for cross checking. Other precautions are taken to afford a fourfold daily verification of loads and charges.

Extended Services

Conversational Use of IMCO Tie Lines

Normally no conversation takes place between a subscriber and his IMCO operator, in completing a call, at either end of the circuit. In fact, were it not for built-in switching disabling means, it would be impossible for a subscriber to ask questions or converse locally with his operator.

The means adopted involves a slow-acting relay[13] which interposes an interval of $1^{1}/_{2}$ seconds between the instant the IMCO operator plugs in and the moment the program switch begins its sweep (see Fig. 3). It will be recalled that in placing a call the subscriber pushes a calling button. His printer motor does not start until the IMCO operator plugs in. Thereafter, for a period of $1^{1}/_{2}$ seconds, the customer is afforded an opportunity to freeze the program switch, simply by touching any key on his keyboard. He may then converse with the IMCO operator at will. Since the program switch has not operated it is not a chargeable call, no bid has been made to London, no trunk has been occupied.

In the reverse sense, the IMCO operator can plug in any sending or receiving cord or his test-printer cord in a tie-line jack and raise the subscriber by bell signals. In this case the regular functions of a plugged-in sending or receiving cord are automatically disabled by a talk-freeze relay (Fig. 3).

Cross Bidding

Cross bidding, in the nature of an extension to usual IMCO connections of the MAMA type described, is provided for in the equipment. By its use, MA New York could ask for occasional connections to MB, QC, or PD London. By adaptation to cross bidding, the IMCO service, in fact, would become a general teleprinter exchange service, because it is in the nature of transatlantic telegraph communication for a customer to call regularly one principal correspondent and others only occasionally.

Cross bids may be placed during the $1^{1}/_{2}$ second interval mentioned, during which the customer may freeze the program switch. If MA wishes to be connected to PD, his "conversation" with the operator consists merely in saying, "PD PLS." The operator pushes his cross-bid key; the program switch steps to point *13;* PDPD is then inserted and the cross-bid key touched again to complete the bid (Fig. 3).

When the cross bid arrives in London, the operator there proceeds exactly as though it were a PD bid, and when he inserts the plug in the PD jack, the usual information is sent back to New York for automatic checking. Thus, cross bids are checked on the overseas bid trunks with the same accuracy as are regular bids.

IMCO Call Simulation

The constant availability of two switching operators with the versatile low-cost reperforator equipment described lends itself to utility uses whether connected or not with IMCO. The overseas trunks assigned to the switched service may be adjusted in number to suit the varying needs of the dispatcher. If their use in non-IMCO service is to be temporary, for example, to set up an overflow circuit to relieve congestion or provide a direct press channel, there is a gain in leaving it in IMCO in order to take advantage of its simple switching facility.

It is obvious that any stop-start termination can be made to simulate an IMCO tie line, be it an operating position in the central office, in a branch office, or on a trunk extension of any length beyond London or New York; and a number of corresponding jacks in the New York and London turrets may be reserved for the making of non-IMCO utility or regular pairings instead of the constant ones under IMCO contract. Of course the traffic so handled would come under appropriate non-IMCO tariffs, and any character counts automatically supplied between the establishment and knockdown of the non-IMCO service would be meaningless for charging purposes.

Non-IMCO Tie-Line Message Switching

A rather large number of non-IMCO tie lines in New York are used for mixed business, domestic and cable, and of the

Vol.
Pt. 1
196

cable usage, a high percentage is for London destination. There is an obvious speed and economic advantage in having such New York–London messages switched direct from the New York customer to a London (rather than a New York) receiving cable operator, and vice versa. Provision is made in the IMCO designs for making such switching transfers.

Such mixed tie lines are multipled into an assigned section of the IMCO turret on jacks identified as to tie-line call letters but unmatched with a calling signal light. When the service is operating, the IMCO operator's attention will be called to the call, by a printer to which the call letters are relayed by one of the regular plan *2* concentrator tie-line operators, on a house pony wire.

Mixed tie lines of this type have double identities. For their domestic and non-London cable business, they use their regular 3-letter tie-line calls and deal with a New York operator. When they want the London operator, they say to the answering tie-line operator: VA PLS, VA being their assigned call for use with London. The tie-line operator removes his plug from the plan *2* concentrator jack, stopping the customer's printer, plugs into the house pony (which contains interference prevention arrangements) and says "VA." The IMCO operator, in response, plugs into the VA multiple and handles the transferred call the same as any other IMCO call, the London-end "subscriber" being the London cable operator. The usual message heading is the calling party's receipt for his transferred London connection. All circuit elements are subject to the IMCO tests for integrity and the IMCO records protect the service.

Such service is also open to cross bidding, in which event all printer services converging in the London office are available for intercommunication to all points served by teleprinter which can be brought into the New York IMCO turret. When this is done a relatively inexpensive general reperforator system is created, having in it only one *FRXD* reperforator, the transmitting feature of which becomes operative before a completed message is in storage. This may be compared with ordinary private-wire reperforator systems, which also connect switching turrets (with their operators) in two cities but which require cross-office transmission with two reperforators at city of origin, a third reperforator at city of destination, and a completed-message requirement on all three reperforators.

Cord Circuit Detail

The foregoing explanations are probably necessary to an understanding of the built-in features of the sending cord circuits of Fig. 3, showing an R cord in the MA jack at New York, connected to send on a varioplex subchannel to London through the *FRXD* reperforator. When plug R is inserted in jack MA, the tie-line circuit from ground is completed through operate relay, connect relay, and setup relay to a vacuum-tube circuit through plan *2* concentrator for a busy test. If clear, negative battery completes the circuit to tie line; tie line closes to marking, applying ground to the connect relay through operate relay contacts to negative at monitor table power relay, starting *FRXD* and operating the connect relay.

The connect relay locks up through setup relay to positive battery: 1. removing busy test; 2. transferring local ground to the *FRXD* transmitting relay; 3. energizing the "go-ahead" point from the program switch preparing it to close the operate relay after the message heading has been run; 4. connecting outbound bid printer into circuit; and 5. operating the release relay.

The setup relay: 1. operates the shake relay to permit the rings of the turret jacks to place call letters on the program switch; and 2. releases the slow-release relay.

After $1^1/_2$ seconds, back contacts of the slow-release relay apply a stepping pulse through the back contacts of the break relay to the stepping magnet of the program switch. The sequence then follows level *1-s*.

After the message heading has been sent, the operate relay receives its pulse from the program switch, which: 1. disconnects the connect relay; 2. locks itself up, maintaining power on the monitor table; 3. applies a locking circuit to the release relay; 4. transfers the coil of the *FRXD* reperforator to marking contact of tie-line relay; and 5. transfers the tie line from the apparatus common to all cords to battery unique to the cord in use.

The customer is now sending, and his tape is being stored. The release relay blocks off the transmitting portion of the *FRXD*. London's confirmation of the bid through the comparator contact puts opposing currents through differential coils. The release relay falls back, removing the block against *FRXD* transmission. The message now goes to London on a varioplex trunk.

The customer's knockdown bell signal on the tie line closes a contact in the *FRXD* receiving element, operating the disconnect relay. The disconnect relay: 1. locks itself up; 2. short-circuits the operate relay and restores it to normal; 3. maintains power on the monitor table to handle any tape remaining in *FRXD;* 4. opens the customer's tie line, shutting down his printer; 5. lights the disconnect signal; and 6. puts ground on *FRXD* to prevent making any false letters when unplugged.

The operator responds to the disconnect signal as soon as the neon light stops flashing showing transmission is ended. The operator unplugs, breaking the circuit through disconnect relay and restoring normality.

If tie-line relay is opened (by the customer) during the interval before the slow-release relay has closed to start the program switch, the break relay operates and locks, preventing the stepping of the program switch.

If a cross bid is made by the customer, the operator throws the cross-bid key momentarily to the nonlocking "down" position, letting the program switch rotate and stop on point *13*. He inserts the cross-bid PDPD on his bid printer and contact *13*. He again throws the cross-bid key down momentarily allowing the complete cross bid to go through. The top (locking) position of the cross-bid switch is used to permit conversation between the operator and the customer on the tie line when the program switch is on point *13*.

The talk-freeze relay works in concert with the tie-line relay as long as the customer does not interrupt the tie line during the interval between plug-up and the time the program switch reaches point *13*. If the customer does interrupt, the talk-freeze relay passes an operating current to the break relay preventing the program switch from starting, or, having started, from going beyond point *13*. This permits the customer to stop or modify a bid.

References

1. A PRINTING TELEGRAPH CONCENTRATOR, William B. Blanton. *AIEE Transactions*, vol. 50, June 1931, pp. 414–25.

2. PUSH-BUTTON SWITCHING, W. B. Blanton, F. L. Currie. *Western Union Technical Review*, New York, N. Y., vol. 1, 1948, pp. 119–27.

3. COMMON SEQUENCING, P. Holcomb, Jr. *U. S. Patent no. 2,291,964*, filed May 13, 1939; granted Aug. 4, 1942.

4. CHARACTER GENERATION IN TELEGRAPH SYSTEMS, F. T. Turner. *Western Union Technical Review*, vol. 4, 1958, pp. 156–60.

5. THE AUTOMATIC TIME AND DATE TRANSMITTER. W. S. W. Edgar, Jr. *Ibid.*, vol. 2, 1948, pp. 150–54.

6. TAPE-TO-PAGE TRANSLATOR, A. E. Frost, *Ibid.*, vol. 3, 1949, pp. 61–67.

7. Switching Telegrams from Tie-Lines to Trunks, F. L. Currie, A. E. Hildreth, Jr. *Ibid.*, vol. 4, 1958, pp. 145–55.

8. A Modern Reperforator Switching System for Patron Telegraph Service, R. F. Dirkes. *Ibid.*, vol. 2, 1948, pp. 138–49.

9. The Development of Western Union Switching Systems, R. F. Blanchard, W. B. Blanton. *Ibid.*, vol. 1, 1958, pp. 2–16, 75–92.

10. The Varioplex—A New Development in Telegraphy, E. R. Shute. *AIEE Transactions (Electrical Engineering)*, vol. 60, Apr. 1941, pp. 181–84.

11. Basic Principles of the Varioplex Telegraph, Philo Holcomb, Jr. *Ibid.*, pp. 1102–07.

12. The Western Union Varioplex Telegraph System, O. E. Pierson. *Electrical Communication*, New York, N. Y., vol. 22, 1944, pp. 101–09.

13. Electronic Timers in Telegraph Switching Systems, E. A. Demonet, L. C. Bower. *Western Union Technical Review*, vol. 4, 1950, pp. 93–100.

◆

Discussion

Austin G. Cooley (Times Facsimile Corporation, New York, N.Y.): This switching system offers many very desirable operating features not found in other arrangements and accomplished with simple standard equipment. The designs take full advantage of the fact that the majority of the connections are established between the same offices.

When other than routine connections are required the switching is accomplished manually. The switching operators can make rapid connections with "insurance" against mistakes, as ingenious safeguards are provided against switching error or connection to an inoperative tie line or called printer. This[9] supervisory method is extremely simple and effective, being of course partly based on the fact that customer *A* in New York normally calls only customer *A* in London, *B* in New York, only *B* in Paris, etc. But provision is made for occasional *A–B* or *B–A* connections by special calling and switching procedure.

Functions are performed automatically on the customer's behalf which usually require him to make accurate and most exacting keyboard manipulation, such as numbering his messages, indicating on the message "origin" and "destination," adding "time" and "date," etc. These functions are performed by automatic equipment (which also includes the testing devices for insuring correct selection of an operative-called station).

While automatic switching gear might be employed, the cost and complexity of such gear is avoided by using a manual cord-type switchboard with automatic alarms and controls to guard against wrong connections. A single "common" control equipment is automatically associated with each cord during the time a connection is being set up and then released for another call.

The day-to-day operation of the Western Union IMCO system affords ample evidence of the value and importance of this development.

Radio-Frequency Interference Considerations in the TD-2 Radio Relay System

H. E. CURTIS
NONMEMBER AIEE

THE NUMBER of microwave systems that may be used on a particular frequency assignment in a given area is determined by the mutual interference between the systems. The Bell System is particularly conscious of this problem because of its extensive use of broadband microwave systems.

At the end of 1958 the Bell System had in operation approximately 215,000 one-way broadband channel miles of microwave systems in the United States, corresponding to approximately 7,000 transmitters and an equal number of receivers. A large portion of this extensive network makes use of the *TD-2* radio relay system[1,2] operating in the 3,700–4,200-mc (megacycles) band. The *TH* system[3] operating in the 6-kmc (kilomegacycles) range is under active development, and the *TJ* system operating in the 11-kmc common carrier band (10,700–11,700 mc) is now in commercial use. Equipments made by other manufacturers also operate in the 4- and 6-kmc bands.

In spite of the highly directive antennas now in use, a certain fraction of the transmitter power from any station will be radiated in directions other than that for which it is intended, and this becomes a potential source of intrasystem interference to itself and also to other systems operating in the same frequency band.

This paper is directed specifically toward the *TD-2* system because of its extensive use and the experience gained with it. However, the philosophy developed herein may be applied directly to the study of interference arising in or from other systems.

It is sufficient for the purposes of this paper to point out that the *TD-2* system, in common with many other microwave systems, uses frequency modulation. It can provide six 2-way broadband channels. Each of these channels can handle two one-way television channels or 480 or more 2-way telephone channels.

The frequency arrangement at a repeater is shown diagrammatically in Fig. 1. The six channels in one direction operate on channel assignments whose midband frequencies are 80 mc apart. The six channels in the other direction are interlaced with the first six, and thus there are 40 mc between midband frequencies of adjacent channels in opposite directions. Thus, the potential interference from a transmitter at a tower is always separated by 40 mc or more from the closest receiving carrier frequency used at the same tower, and r-f (radio-frequency) and i-f (intermediate-frequency) filtering is sufficient to reduce such interference to a negligible magnitude.

When a 2-way channel such as that shown in Fig. 1 is extended to include three consecutive repeaters, as in Fig. 2, it will be noted that two additional types of interference are involved. Similar interferences will exist, of course, on the other five channels.

For convenience in discussing these interference paths, let each antenna involved be given a number followed by a pair of letters designating the direction of transmission and whether the antenna is transmitting or receiving. Thus the sequence from west to east is $1E(T)$ to $2W(R)$ to $2E(T)$ to $3W(R)$, etc. At any particular receiving antenna such as $2W(R)$ there will be normally two important interferences. One is from $1W(T)$ into $2W(R)$, and the second is from $3W(T)$

Paper 60-54, recommended by the AIEE Communication Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., January 31–February 5, 1960. Manuscript submitted October 22, 1959; made available for printing November 23, 1959.

H. E. Curtis is with the Bell Telephone Laboratories, Inc., Murray Hill, N. J.

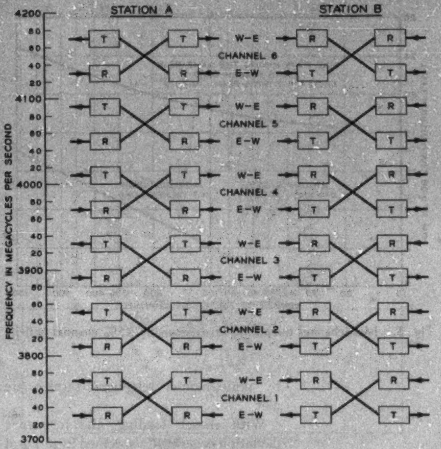

STATION A
STATION B
FREQUENCY IN MEGACYCLES PER SECOND
4200
80
60
40
20
4100
80
60
40
20
4000
80
60
40
20
3900
80
60
40
20
3800
80
60
40
20
3700
T
R
W-E
E-W
CHANNEL 6
CHANNEL 5
CHANNEL 4
CHANNEL 3
CHANNEL 2
CHANNEL 1

Fig. 1. Frequency arrangement at repeater

T = Transmitter
R = Receiver

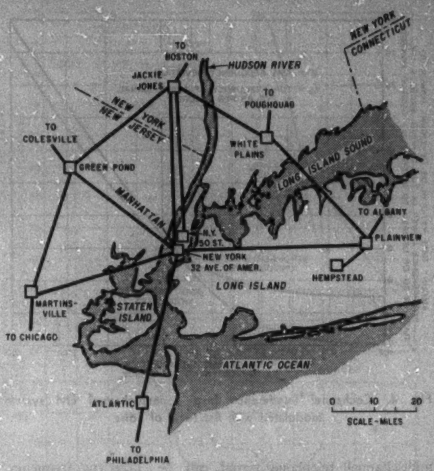

TO BOSTON
HUDSON RIVER
NEW YORK
CONNECTICUT
JACKIE JONES
TO POUGHQUAG
NEW YORK
NEW JERSEY
TO COLESVILLE
WHITE PLAINS
GREEN POND
LONG ISLAND SOUND
MANHATTAN
TO ALBANY
PLAINVIEW
N.Y. 50 ST.
NEW YORK 32 AVE. OF AMER.
HEMPSTEAD
LONG ISLAND
MARTINS-VILLE
STATEN ISLAND
TO CHICAGO
ATLANTIC OCEAN
ATLANTIC
0 10 20
SCALE-MILES
TO PHILADELPHIA

Fig. 3. Microwave systems in New York area

into $2W(R)$. The first, for convenience, is designated "same section," since it transverses the same repeater section as the desired carrier from $1E(T)$; the second, $3W(T)$ into $2W(R)$, may be designated "adjacent section" for obvious reasons. Each antenna is normally subject to these same two interferences.

Of particular significance is the fact that the interference has the same nominal frequency as the carrier with which it interferes, and, hence, it cannot be discriminated against by filters. Therefore, the only protection lies in making the backward response of the antennas adequately low.

The interference situation becomes increasingly complex at a point where a number of converging microwave systems operate on the same or nearly the same frequency. An extreme example of such a point is that which exists at New York, where the Bell System has two centers from which microwave systems radiate, as shown in Fig. 3. Here we have a total of five converging routes with angles ranging from 50–110 degrees, with a sixth paralleling route.

Relation Between R-F Interference and Baseband Interference

The relation between radio interference and baseband interference depends, among other things, on the kind of baseband signal being transmitted. In general, in Bell System use, this is either a video signal or a number of telephone channels arranged in frequency division. The latter case is emphasized herein, since it has been found to lead to a philosophy that provides adequate freedom from interference when either telephone or television signals are transmitted. For analytical and experimental purposes a

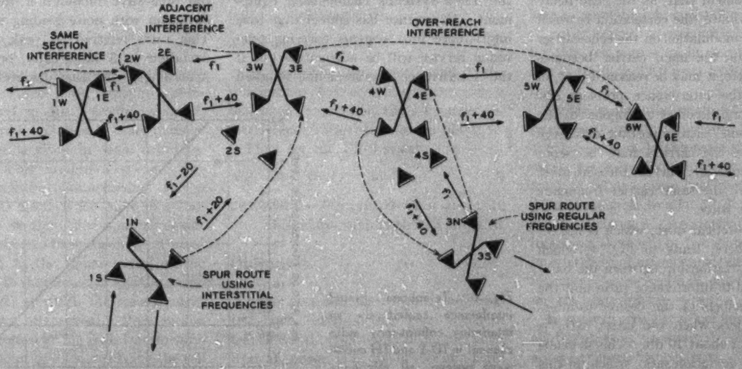

ADJACENT SECTION INTERFERENCE
SAME SECTION INTERFERENCE
OVER-REACH INTERFERENCE
f_1
f_1+40
1W
1E
2W
2E
3W
3E
4W
4E
5W
5E
6W
6E
2S
4S
1N
1S
3N
3S
f_1-20
f_1+20
f_1+40
f_1
SPUR ROUTE USING REGULAR FREQUENCIES
SPUR ROUTE USING INTERSTITIAL FREQUENCIES

Fig. 2. R-f interferences on the TD-2 microwave system with typical paths as shown

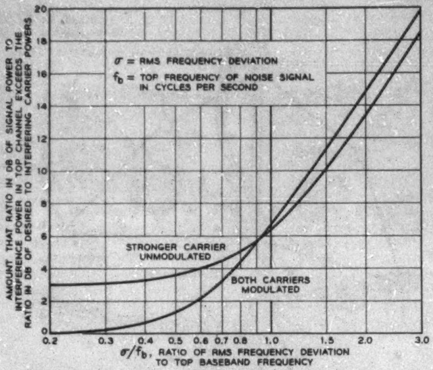


Fig. 4. Cochannel interference from a multichannel FM system modulated with flat band of noise

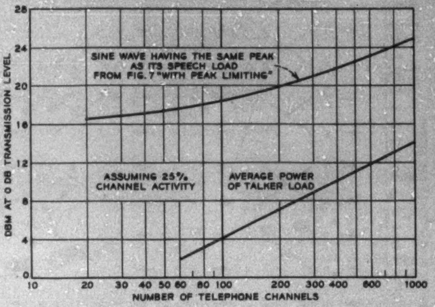


Fig. 5. Multichannel telephone load, assuming 25% channel activity

multichannel telephone signal can be adequately simulated by a band of random noise of the same average power and the same bandwidth as the multichannel signal.

The problem of intermodulation due to echoes in FM (frequency-modulation) with noise loading has been treated in a previous paper,[4] which showed that the amount of intermodulation depends, among other factors, upon the echo delay. (During the preparation of this paper, a theoretical paper[5] on the interference between noise-modulated FM carriers appeared, which gives a general expression for the interferences as a function of frequency deviation, top baseband frequency and frequency separation between carriers. The resulting equation is identical with the one from which Fig. 4 was computed.) It is of significance to the present problem that, as the echo delay becomes infinite, the correlation between the signal modulation on the echo carrier and that on the main carrier becomes zero. Hence, it may be reasoned that, in the limit, the interference problem discussed herein becomes analogous to the echo problem. An important difference is that here the carriers may be separated in frequency, whereas the echo must always have the same carrier frequency as the main wave.

This theoretical work, which is not reproduced here, leads to Fig. 4, which shows the relationship between the baseband signal-to-interference ratio and the ratio in decibels of the desired-to-interfering carriers when the latter ratio is greater than about 10 db. This depends on the FM deviation ratio, which, in this case, is specified in terms of the ratio of the rms frequency deviation to the top baseband frequency. Furthermore, the relationship depends somewhat on whether both carriers are modulated with noise or only the weaker is so modulated. In the latter case, the baseband signal to which the interference is referred is assumed to be equal in magnitude to that applied to the interfering carrier.

In addition to the interference from the modulation on the carriers evaluated in Fig. 4, there is generated a beat note having a baseband frequency equal to the difference in frequency between the desired and interfering carriers. Since the beating oscillator frequencies in the *TD-2* system are crystal-controlled, the carrier differences are held to within a few tens of kilocycles. Hence, the interferences between the carriers themselves fall below the lowest frequency used by the *TD-2* system's channelizing equipment. Experience has shown that tone interference into systems carrying television service will be adequately low if the interference requirements imposed by multichannel telephone service are met.

With speech loading, the frequency deviation is generally specified in terms of the peak value in contrast to noise loading where it is convenient to refer to it in terms of the rms frequency deviation. Fig. 5 shows the sine wave power, at a point of 0-db transmission level, which has the same peaks as a multichannel telephone system, assuming an activity factor of one-quarter. This is taken from a paper by Holbrook and Dixon.[6] Also shown in Fig. 5 is the power of a multichannel load, based on talkers whose volumes correspond to the mean power of the assumed talker distribution curve, an activity factor of one-fourth being included. This curve is derived from constants given in reference 6.

The *TD-2* system with a 480-channel load normally uses a peak deviation with speech of ±4 mc. Fig. 5 shows that for this channel load the ratio of the rms frequency deviation with equivalent full sine-wave modulation to the rms deviation with noise loading is 11.5 db. This ratio referred to peak sine-wave deviation is 3 db greater, or 14.5 db. Accordingly, the rms deviation with a

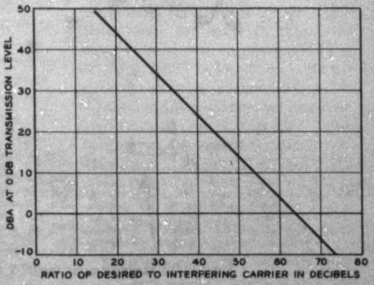


Fig. 6. Telephone channel interference caused by an interfering cofrequency radio channel in TD-2 and TH microwave systems, all telephone case

Vol.
Pt. 1
196

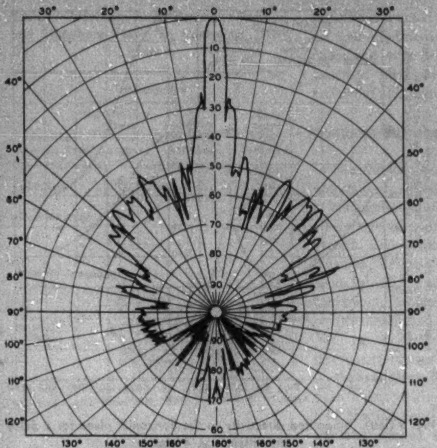


Fig. 7. Measured directivity pattern of an individual antenna

Horn reflector
Horizontal directivity
Vertical polarization
3,740 mc
3-db beamwidth = 2.5 degrees

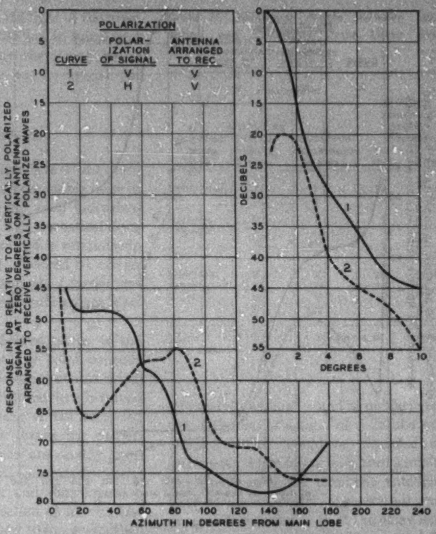


Fig. 8. Smoothed directivity in the horizontal plane for a horn-reflector antenna at 4 kmc

noise load can be seen to be 0.7 mc. Therefore, since the top baseband frequency is approximately 2 mc in the example, the ratio of rms frequency deviation to top baseband frequency would be approximately 0.35. Hence, Fig. 4 shows that, in this case, the baseband signal-to-interference ratio in a telephone channel at the top baseband frequency is 1/2 db greater than the ratio in decibels of the "desired" carrier to an interfering carrier when both carriers are modulated with noise. Thus, by rounding this value to 1 db, we may write in db

$$S_{\text{dbm}}-N_{\text{dbm}}=C-I+1 \qquad (1)$$

It has been found useful in the Bell System to express noise power in a telephone circuit in dba. (A discussion of this unit and its relation to other units of noise power is given in some detail by Franke.[7]) This implies a specific frequency weighting characteristic, and 1 milliwatt (0 dbm) of flat noise in a 3-kc band reads 82 dba. For a measurement of noise power to have meaning, it must be referred to the signal power at the same point in the circuit, or else the transmission level at which the measurement is made must be specified. Noise in dba is often specified at the 0-db transmission level, the point of reference being the toll transmitting switchboard.

Holbrook and Dixon give the power of an rms talker at a point of 0-db transmission level (0-db TLP) as −9.9 dbm. With a channel activity factor of one-quarter, the equivalent noise load per 4-kc channel is −15.9 dbm. The noise in a 3-kc band is 1.2 db less, or −17.1 dbm. This amount of noise power would measure very nearly 65 dba. From this it follows that, given a measurement of signal power and of noise power in the same narrow band, the equivalent noise in dba at a point of zero transmission level can be obtained from the following relationship:

$$N_{\text{dba}}=65-S_{\text{dbm}}+N_{\text{dbm}} \qquad (2)$$

By combining equations 1 and 2, the relationship between interference in a telephone circuit and the r-f carrier-to-interference ratio can also be obtained. Thus, for the *TD-2* system

$$N_{\text{dba}}=64-(C-I) \quad \text{in db,} \qquad (3)$$

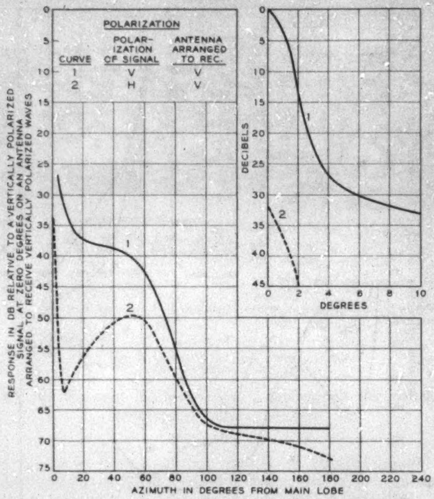
provided $C-I$ is greater than about 10 db. This relationship is plotted in Fig. 6.

Objectives

Experience has shown that long-haul telephone circuits will give excellent transmission performance if the rms noise from all sources is 38 dba or less at a point of zero transmission level (0-db TLP). On a radio relay circuit, noise in the more general sense may arise from many sources. A portion will be in the form of fluctuation noise arising in the converters; some will be intermodulation because of amplitude and phase distortion in the i-f and r-f circuits, as well as from echoes in the wave guide runs and the interconnecting i-f cables. It is necessary to divide the total permissible noise among all the known contributors, allotting to each source an objective consistent with its importance and the feasibility of meeting it. Of particular concern here is unintelligible crosstalk caused by r-f interference; a reasonable allocation for this last source of impairment is 24 dba at the 0-db TLP for a 4,000-mile system.

Provision must be made for both the interference within the main route itself and interference to the main route from other systems that converge. Of the total of 24 dba given previously, it appears reasonable to allot 22 dba to main route interference. As mentioned in the foregoing, there will be two main-route r-f interferences of major importance at each repeater: "same section" and "adjacent section" interference.

In order to arrive at a "per interference" allotment, it is necessary to discuss the antenna directivity patterns. The measured directivity pattern of an individual



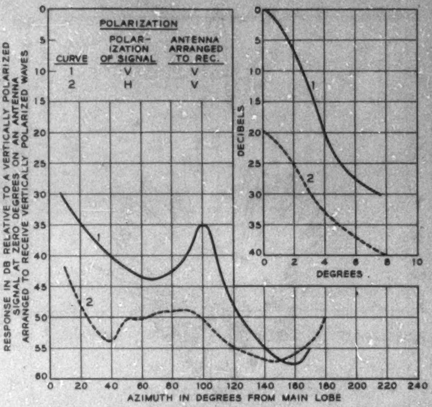


Fig. 9 (left). Smoothed directivity in the horizontal plane for a delay lens antenna at 4 kmc

Fig. 10 (above). Smoothed directivity in the horizontal plane for an 8-foot parabolic antenna at 4 kmc

sample of a given antenna at a particular frequency will exhibit numerous maxima and minima, as shown by Fig. 7. It is not practical to make use of the sharp nulls to reduce converging route interference, because their angular position varies with frequency and is apt to differ from one particular antenna to another. Also, the nulls are usually sharper than the angular stability of the tower used to support the antenna. Therefore it is desirable to use smoothed directivity patterns, as shown in Figs. 8, 9, and 10, in designing particular systems. The smoothed patterns were obtained by drawing a more or less smooth curve through the peak (or poorest) values of the measured patterns. Therefore, these effectively are "worst" values rather than "rms" values.

In a multirepeatered system where there is a sizable number of exposures involving random angles, a substantial percentage of these exposures will contribute less interference than would be indicated by the work presented here. Therefore, the total rms interference is somewhat less than would be expected by simple power addition of the individual interferences. An examination of various measured directivity patterns indicates that it is reasonable to assume that straight power addition is perhaps too conservative by about 6 db. This factor can be introduced into the allotment either by altering the antenna characteristics or by altering the allotment. Since the antenna characteristics describe what was actually measured, we have chosen not to modify them; instead, we have modified the allotment.

Furthermore, considering one exposure alone, intelligible crosstalk can appear only if the two carriers involved are within a few hundred cycles of each other; the chance of this is very small. Actually, the two carriers may be as much as several tens of kilocycles apart in frequency; hence, an observer would more probably hear crosstalk from two telephone channels simultaneously with a random amount of frequency staggering. Since the frequency differences between carriers in a large number of exposures will be quite randomly distributed about an average, the telephone interference will appear as babble and may be treated as noise. Therefore, on the basis of 140 repeaters (4,000 miles), two interferences per repeater, and a factor of 6 db as previously mentioned, the per repeater allotment in dba for either the same section or the adjacent-section interference is

$+22-10\log 280+6$
or approximately $+4$ dba at 0-db TLP

Since this allotment is slightly below that for first circuit noise, the sideband power cycle of allowable r-f interference is slightly below the fluctuation noise per cycle in the receiver converter in which the r-f interference falls.

It is evident then that during a fade of the desired carrier entering any particular receiver converter, the converter noise and the r-f interference contributions, as observed in any telephone channel at the terminal of the system, will rise together and always stay in balance. Obviously, if the interfering carrier fades, it becomes increasingly weaker than the fluctuation noise, and the effect then becomes entirely negligible.

Pertinent features of a long main route such as may extend across the continent are the spurs by means of which television and telephone channels are brought to cities situated off the main route.

At each point where a spur joins a main route four r-f interferences will occur, as shown in Fig. 2. Two of these will appear finally as telephone channel babble at the west end of the backbone route, and similarly two will appear at the east end. Thus, assuming 64 spurs per 4,000 miles, there will be a maximum of 128 such interferences at either end.

However, the angles of the spurs may be expected to be distributed at random, with presumably no angle less than a value consistent with the objectives developed below. Hence, many interferences may be expected to contribute substantially less baseband noise than the alloted value. In view of this it is probably not unreasonable to assume that only one-half of these interferences will contribute significantly to the total.

The maximum amount of baseband noise each of these r-f interferences may be

Table I. Minimum Angle Between Disturbed and Disturbing Paths for Cochannel Interference

	Polarization		
Disturbed Antenna Type	Disturbing (Transmitting) Antenna	Disturbed (Receiving) Antenna	Minimum Angle, Degrees
Delay lens	V	V	80
	H	V	72
Horn-reflector	V	V	58
	H	H	26
	H	V	10
	V	H	11
Parabolic, 8-foot	V	V	144
	H	H	X*
	H	V	130
	V	H	26
Parabolic, 10-foot	V	V	122
	H	H	X
	H	V	110
	V	H	X

* X indicates that the objective cannot be met at any angle with these conditions.

permitted to contribute can then be obtained by using the same line of reasoning as was followed above. Thus, assuming 64 spurs per 4000 miles, the per spur allotment is

$$+20-10\log 64+6$$
or $+8$ dba at 0-db TLP

It is recognized that the foregoing, at best, can be only a guide, since any tentative layout may call for spur angles ranging from zero (which may introduce excessive interference) to 180 degrees. Hence, each specific spur contribution must be examined in the light of the overall spur allotment, bearing in mind that cases will arise where there may be no practical alternative but to exceed the pro-rated allotment.

Systems Application

The previous sections have 1. set down objectives in terms of permissible noise in a telephone channel for each individual r-f interference on a hypothetical long system, and 2. developed a relationship between the ratio of a desired r-f carrier to an interfering cochannel r-f carrier and the telephone channel interference that results therefrom.

The ratio of the desired carrier to the interfering carrier at any point depends on the transmitted powers, antenna gains, path losses of the two carriers involved and the discrimination of the receiving antenna against the interfering carrier. The first three factors are readily computed and need not be elaborated on here. The antenna's gain and discrimination are, of course, a function of its size and design. Figs. 8, 9, and 10 show smoothed discrimination curves for the antennas normally used in the *TD-2* system: the horn-reflector antenna, the delay lens antenna and a simple 8-foot parabolic antenna.

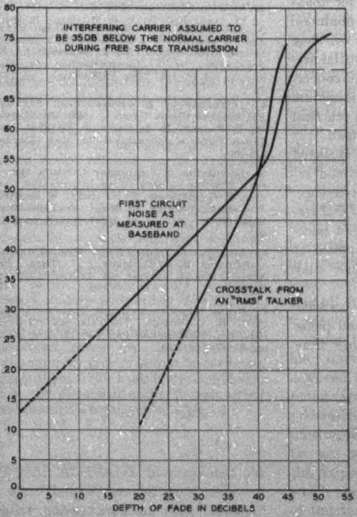


Fig. 11. First circuit noise and adjacent channel crosstalk with interfering carrier assumed to be 35 db below the normal carrier during free-space transmission

Table II. Minimum Angle Between Disturbed and Disturbing Paths Using Slot Frequencies

	Polarization		IF Filter, Minimum Angle, Degrees	
Disturbed Attenna Type	Disturbing (Transmitting) Atenna	Disturbed (Receiving) Antenna	Out	In
Delay lens	V	V	12	3.5
	H	V	*	*
Horn-reflector	V	V	6	3.1
	H	H	9	5.5
	H	V	4	2.5
	V	H	6	2.4
Parabolic, 8-foot	V	V	20	5.0
	H	H	20	7.2
	H	V	5	1.8
	V	H	8	4.9

* Less than 2 degrees; a minimum of 2 degrees is suggested.

Since the response of the 8-foot parabolic antenna to a signal from the rear is only 45 db below the main lobe response, both the "adjacent section" and "same section interference" would be on the average +19 dba, or a total of +22 dba per repeater. This would be regarded as intolerably high for a long-haul circuit. For this reason, the parabolic antennas have been used mostly on short-haul spur or secondary routes. Both the delay lens and horn-reflector provide in excess of 65 db discrimination in the rearward direction, which is entirely adequate even for systems of 4,000 miles in length.

The minimum angle between a spur route and a main route on the basis of a given permissible interference is determined by the discrimination of the antennas involved. This, for a given pair of antennas, depends on whether the interference is polarized similarly or at right angles to the plane of polarization of the desired carrier.

Table I gives the minimum angle between the disturbed and disturbing paths based on the interference into the receiving antennas at the point of convergence, with the inherent assumption that the disturbed and disturbing signals are equal in strength; i.e., they originate from similar equipments and traverse paths of equal length.

In general, the Bell System does not operate separate microwave routes in close enough proximity for mutual interference to become a consideration. However, where it is suspected, the baseband interference can be computed on the basis of the principles laid down above. In general, disturbing transmitters and the disturbed receivers on separate routes do not point at one another, and the discrimination advantage of both antennas is obtained. The combinations and possibilities are so great for such cases that generalization becomes useless and each case should be treated individually.

In the standard frequency arrangement

for the *TD-2* system, transmitting and receiving carriers are separated by 40 mc. Ahead of the receiver converter, r-f filters provide 20 db of discrimination to the transmitter. This and the i-f selectivity are more than adequate to eliminate interference between transmitters and receivers at the same site.

It is possible under certain conditions to operate channels in the 20-mc slots between an adjacent transmitter and receiver in the standard frequency arrangement. These interstitial or so-called "slot" channels are used at converging points or spurs where co-channel interference considerations would limit the minimum angle of convergence to an undesirably large angle. (Developments subsequent to the preparation of this paper indicate that additional discrimination applied at each repeater at the intermediate frequency of 70 mc reduces the crosstalk from the adjacent microwave channels 20 mc offside so as to allow the minimum angle between converging routes to become zero degrees. This is equivalent to saying that six interstitial 2-way microwave channels can also be operated in the 3700- to 4200-mc band, making a total of 12 broadband channels. The technical problems relating to the possible use of these six additional 2-way channels are now being studied at Bell Telephone Laboratories.)

When interstitial frequencies are used each channel may have interferences separated only 20 mc from the normal carrier frequency, i.e., at 50 mc and 90 mc in the i-f band. This separation is not sufficient to provide any appreciable suppression of the interfering carriers by r-f filtering. Furthermore, the i-f discrimination to such interferences is only of the order of 5–10 db.

If the FM receiver could be assumed to consist of a perfect limiter and a perfect demodulator, FM theory would indicate that the resulting interference at baseband frequencies would be infinitesimal in magnitude in the frequency range occupied by the normal signal. Furthermore, any such interference would be incoherent with the modulating signal on the disturbing channel. Actually, experiment has shown that the signal on the weaker carrier is transferred to the stronger carrier. The crosstalk increases 2 db as the ratio between the desired carrier and the interfering carrier decreases 1 db, provided that the interferring carrier is weaker than the desired carrier; beyond this point, the baseband interference rises still more rapidly. No completely satisfactory explanation has been advanced, other than the generalized one that physical amplifiers, limiters, and discriminators are frequently not ideal devices.

From a systems engineering viewpoint, when interstitial frequencies are used the interfering carrier must be sufficiently weaker than the desired carrier during free space transmission that, when the desired carrier fades, the crosstalk remains below the fluctuation noise as long as the circuit is useful.

Fig. 11 shows how the fluctuation noise theoretically increases with depth of fade on a single typical repeater section. Also in Fig. 11 is a curve showing how the crosstalk from an interstitial channel increases with a fade, assuming that the interfering carrier is 35 db weaker than the interfered-with carrier during periods of free space transmission. It is assumed, in this case, that the interfering channel is modulated with average power talkers.

This figure indicates that, with a carrier ratio of 35 db, the crosstalk remains below the noise for fades up to about 35 db. Arrangements are provided automatically to switch the faded channel to a protection channel when the depth of fade exceeds 35 db.

Since this carrier ratio is about 20 db less than the requirement for tolerable cochannel interference, the minimum angle at which two routes may converge is greatly reduced when interstitial frequencies are used at the point of convergence. Table II illustrates the minimum angles of convergence that are practical when interstitial frequencies are used.

In a linear system it would be expected that additional i-f selectivity would reduce the effect of interstitial interference. Experiment has shown that the *TD-2* system is so nonlinear that added i-f selectivity is effective only when applied at each repeater where interference may enter the system.

For this purpose i-f filters are available that have no appreciable loss and phase distortion in the frequency range from about 60–80 mc and about 30 db loss in the frequency regions of 50 and 90 mc. When these filters are used in a *TD-2* system using an interstitial frequency plan, the angles between converging routes need be such as to provide an antenna discrimination of only 25 db during periods of free-space transmission. Table II gives the minimum angles necessary for adequate protection from interference when i-f filters are used.

When the suggested minimum angles are very small, as when filters are used, allowance must be made for the mechanical stability of the towers and for the precision to which the antennas can be oriented. The former depends on the design of the tower structure and on the loads to which it is subjected. The latter depends on the method and skill used in orienting the antennas. No safety factors for these considerations have been included in the suggested minimum angles.

Also for these same reasons, with the delay lens antenna and a crosspolarized interfering carrier, it is suggested that the minimum permissible angle between converging routes be 2 degrees rather than 0 degrees as would be indicated by the discrimination pattern for the delay lens antenna.

References

1. A Broad-Band Transcontinental Radio Relay System, **T. J. Grieser, A. C. Peterson.** *Electrical Engineering*, vol. 70, Sept. 1951, pp. 810–15.

2. The TD-2 Microwave Radio Relay System, **A. A. Roetken, K. D. Smith, R. W. Friis.** *Bell System Technical Journal*, New York, N. Y., vol. 30, pt. 2, Oct. 1951, pp. 1041–77.

3. 6,000-Megacycle-Per-Second Radio Relay System for Broad-Band Long-Haul Service in the Bell System, **M. B. McDavitt.** *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 76, 1957 (Jan. 1958 section), pp. 715–22.

4. Interchannel Interference in FM and PM Systems Under Noise Loading Conditions, **W. R. Bennett, H. E. Curtis, S. O. Rice.** *Bell System Technical Journal*, vol. 34, May 1955, pp. 601–36.

5. Distortion in Frequency Division Multiplex FM System Due to an Interfering Carrier, **R. G. Medhurst, E. M. Hicks, W. Grossett.** *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 105, pt. B, May 1958, pp. 282–92.

6. Load Rating Theory for Multi-Channel Amplifiers, **B. D. Holbrook, J. T. Dixon.** *Bell System Technical Journal*, vol. 18, Oct. 1939, pp. 624–44.

7. Noise Measurements on Telephone Circuits. **H. C. Franke.** *Tele-Tech and Electronic Industries*, Philadelphia, Pa., vol. 14, Mar. 1955, pp. 85–87, 152, 153.

Optimum Design Considerations for Radio Relays Utilizing the Tropospheric Scatter Mode of Propagation

CHARLES A. PARRY
MEMBER AIEE

Synopsis: There are many areas of the world where the service which can be provided by the scatter link is needed, but cannot be justified economically. One approach toward reducing costs, suggested here, is to investigate the improvements achieved through optimizing system-design parameters.

It is shown that the nonlinear noise introduced by truncating the bandwidth may be used as a basic tool in this analysis. It is found that the signal-to-thermal-noise ratio may then be expressed purely as a function of traffic volume and noise bandwidth. Data are given relating this to such factors as trunk occupancy, speech volumes, grade of service offered by trunks, and the percentage of time during the busy hour for which the given performance is offered.

This approach leads to a minimum permissible noise bandwidth B given by $B = 15.2\ N$ kc, where N is the number of contiguous 4-kc channels extending from zero frequency. General expressions for the optimum bandwidth as a function of signal-to-noise ratio and traffic loading are also developed. From these expressions it is possible to derive the minimum permissible transmitter power for the optimized system. When related to performance over a smooth spherical earth, it is concluded there is a wide field of application for 100- and 250-watt installations. It is suggested that the optimum design (propagation) reliability should be established from the slope of the bivariate probability distribution for instantaneous fading; this leads to a limiting value of 99.9% on a yearly basis.

Expressions for minimum power also leads a design efficiency factor, expressed as the ratio of optimal power to actual power. In addition, an economic index is discussed; this permits a comparison between various links carrying different numbers of channels and with different grades of service and reliability. Methods of reducing transmitter power by stabilizing some of the parameters are reviewed; a new technique is described in which the bandwidth is controlled by traffic volume, and it is shown that further reductions in transmitter power are thereby possible. A means for the greater utilization of circuits carrying light traffic loads is also considered. From the discussion outlined, commercial breakthrough is considered feasible.

IT HAS ALREADY been pointed out that the commercial application of scatter poses something of a paradox.[1] The physical conditions which tend to lead to the use of scatter, in themselves, imply a social independence between the connected groups which precludes economic justification for the link. Even in strategic applications where normal economic evaluation cannot be applied, the scope of services which can be provided must conform to national budgetary allocations. Under these circumstances even strategic planners must carefully evaluate a communication technique which may cost as high as $10,000 per route mile, and require 250-watt transmitter power per decibel (db) signal-to-noise ratio.

Some of the factors which enter into economic justification have been discussed[1] and a technique for this evaluation in specific terms has been outlined.[2,3] It is apparent that there are many factors involved beyond basic costs alone; nevertheless, any possibility of extensive commercial utilization must be approached by considering the ways in which these costs may be reduced.

Certainly there have been developments in methods of fabrication which have improved the economic picture, but despite this, over-all costs associated with the scatter system remain formidable. When assessed in terms of available revenue, there are many areas of the world where such systems are technically desirable but economically unsound.

Undoubtedly new techniques, for instance controlled carrier operation[4] and low noise preamplifiers, will prove useful in specific applications; however, in many cases, the approach toward minimal over-all cost systems tends to offset advantages so gained.

So far, little has appeared on the possible improvements which might be achieved by optimizing the normal system-design parameters. While a comprehensive analysis of all the factors which must apply to the design of the radio relay are beyond the present scope, it is the purpose of this paper to discuss those which appear most significant. Apropos of this it was pointed out in a previous discussion, that the optimum FM (frequency modulation) system would be "peak" limited; that is, one where the per-channel deviation would be established by the "overload point" of the system.[5] Although this parameter is not normally used in FM-system design, the concept will be developed in the following text.

This discussion is restricted to FM relays using FD (frequency division) channeling equipment.

Initial Approach

The basic starting point for the analysis of communication circuits is the signal-to-noise ratio obtained in the information bandwidth.[6-8] The total circuit noise in practical systems is comprised of nonlinear and thermal components. The calculation of the principal component of the thermal noise which appears in the channels of the multichannel system has been treated extensively and will not be detailed here.[9]

The signal-to-weighted-thermal-noise ratio, S_t, in the top channel of the (single-link) FM system, with the r-f (radio-frequency) signal power equal to the FM threshold, can be expressed in the following manner:

$$S_t = \frac{K'\delta B_\epsilon}{2b\ T_e}\left[\frac{\Delta}{f_m}\right]^2 \tag{1}$$

The threshold power P_t for which this applies is given by:

$$P_t = KTFB_\epsilon \tag{2}$$

The terms in these expressions are defined as follows:

K' = weighting factor. For 4-kc spacing between channels, $K' \simeq 2$.

δ = improvement due to the use of pre-emphasis and de-emphasis networks. This will depend on the particular networks employed; for typical conditions $\delta = 2$ in the top channel.[10]

B = noise bandwidth cycles. For all practical purposes this is the 3-db i-f (intermediate-frequency) bandwidth.

ϵ = minimum permissible ratio of r-f signal power to (rms) noise power in the band B and referred to receiver input. This is defined sometimes as the FM threshold factor.

b = channel spacing in cycles. It is considered to be small compared to the absolute base-band frequency corresponding to the center of the channel.

Paper 60-7, recommended by the AIEE Radio Communication Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., January 31–February 5, 1960. Manuscript submitted November 3, 1958; made available for printing December 4, 1959.

CHARLES A. PARRY is with Page Communications Engineers, Inc., Washington, D. C.

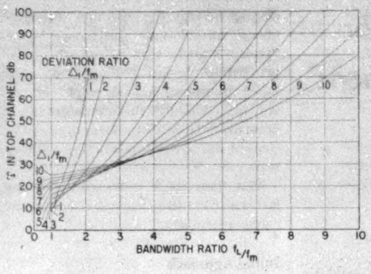


Fig. 1. Signal-to-noise ratio γ measured in top channel of system with white-noise signal extending from $0 \rightarrow f_m$. $2f_1$, truncated bandwidth, Δ_1 peak deviation

Δ = "peak" deviation obtained with a sine-wave signal having an rms power equal to that of the mean power of the multichannel signal (measured in the base band). $\Delta = 1.4 \times$ rms deviation cycles

f_m = modulating frequency for the top channel cycles.

T_c^{-1} = ratio of single-channel test-tone power to mean power of the multichannel (base-band) signal.

F = effective noise figure of receiver.

If the factors of equation 1 were unrestricted, system design would present very little problem, since Δ could be suitably increased for any finite B. However, in practice, as is well known, transmission of the complex wave comprising the multichannel signal, through distorting elements in the system, introduces nonlinear noise into the channels,[9,11,12] and this increases with Δ. A substantial part of the total nonlinear noise tends to vary inversely with B. Equation 2 indicates that for small P_t, B should also be small. Thus, the possibility of a compromise value between Δ and B is indicated. Values ascribed to Δ and B in the literature vary considerably, although the CCIR (International Radio Consultative Committee)[13] has now stipulated per-channel deviation for international circuits. As far as is known, no attempt to correlate Δ and B has been made.

Nonlinear noise associated with B arises from two principal sources. The first is due to loss of side-band energy since B is finite, and for the FM wave, a theoretically infinite bandwidth is required. This is the "truncated bandwidth" effect of Medhurst, and means for calculating this have been derived.[14] This truncated bandwidth may be taken as equal to the noise bandwidth in practical designs. The second source of noise is associated with phase nonlinearity due to the passive networks which determine B. This latter problem has been analyzed by a number of workers,[15—24] and reviews of the principal techniques used have been given by Brown,[16] and Collings and Swirzynaski.[17] Unfortunately, the assumptions which must be made to arrive at workable solutions, tend to reduce the validity of the end result. Most authors use a quasi-stationary format, and Franz indicates certain conditions under which this is valid for the dynamic condition.[18] However, while reasonable agreement can be obtained for simple circuits, there are large divergencies among authors for the general case. Brown, in fact, concludes that a satisfactory analytical solution is not possible. Collings,[17] in deriving coefficients for a typical coupling circuit, indicates that difficulties of solution limit the magnitude of the parameters which could be used for this purpose. On the basis of this work, it does not appear reasonable to derive a general relation for the nonlinear noise power in terms of bandwidth and peak deviation and the phase derivative.

When large numbers of networks in tandem are contributing more or less equally to the total distortion, the resultant bandwidth is substantially greater than that pertaining to a single network with the same total distortion. Typical interstage couplings, such as critically coupled double-tuned[19] circuits, maximally flat triples,[17] and compensated single- and double-tuned circuits,[25] can be designed with acceptable values of distortion. The noise bandwidth of typical interstages is virtually equal to the 3-db bandwidth[26] so that this single stage will also determine B.

Thus it would seem feasible to use a single stage, or at most, a minimal number of stages, in order to control B and at the same time achieve satisfactory phase linearity. The principal nonlinear noise components arising from this source vary at most as the third order of B; on the other hand, the components arising from the truncation effect will vary as the twelfth to thirtieth order.

Hence, if B is varied, this latter effect rapidly becomes controlling and is equivalent to behavior at overload in an AM (amplitude modulation) system. It is feasible to introduce the nonlinear noise from this source as a limiting factor in the design, since the system may then be regarded as "peak limited."[6] This establishes the optimum condition since it per-

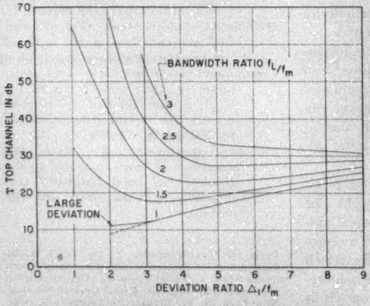
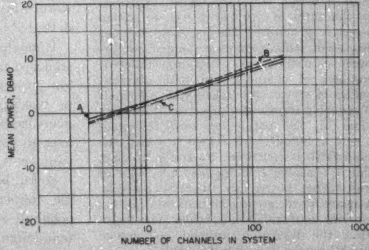


Fig. 2. Fig. 1 plotted to show limiting effect of excessive deviation. Symbols as in Fig. 1. Large deviation curve obtained from Medhurst's "large deviation" formula

Fig. 3 (right). Mean power of multichannel signal; ordinate value not expected to be exceeded for more than 1% of busy hour. All curves are for uncontrolled volumes without limiting



A—From Holbrook and Dixon, reference 27, and Fig. 6
B—From reference 8, volume of single talker equivalent to −15 dbm 0. Spurious signal of −20 dbm 0 per channel included, to simulate carrier leak and signaling tones
C—As for B, but without allowance for spurious signal

Vol.
Pt. I
196

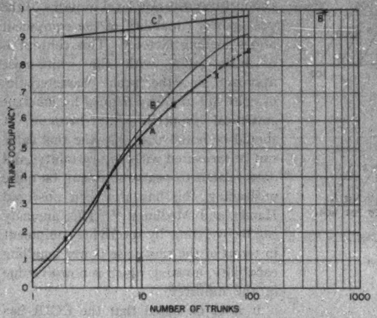


Fig. 4. Trunk occupancy derived from several sources. X points based on Erlang's formula, with full availability, 5% grade of service

A—Operator delayed calls. 5% probability of congestion; reference 33

B—Pure chance traffic, full availability; grade of service, 0.033; reference 33

C—CCIF, 20% deferred calls, 30% failing; reference 36

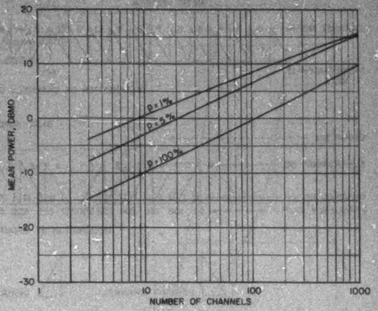


Fig. 5 (right). Mean power of multichannel signal during busy hour based on trunk occupancy for 5% grade of service, or 5% probability of congestion with manual boards

p = probability that ordinate value will be exceeded

p=100% curve obtained from residual effect of carrier leaks and signaling tones

mits the minimum transmitter power for a given set of performance criteria. Hitherto, this approach does not appear to have been used. It will be developed further in a later section of this paper.

Fig. 1 shows the signal-to-noise ratio γ calculated for the top channel of the multichannel system when using a white-noise test signal. This is based on Medhurst's equations 7 and 8. The transition from his "large" to "small" deviation formulas is obtained by replotting the data in the form shown in Fig. 2. In Figs. 1 and 2, Δ_1 is equal to Medhurst's term $(S/2\pi)$.

Mean Power of Multichannel Signal

Since T_c is associated with single-channel test tone, it is numerically equal to the mean power of the composite multichannel signal in dbm 0 (db referred to 1 milliwatt at a point of zero level). This is determined by the statistical properties of speech as well as the number of channels; system design is usually based on that level which will not be exceeded for more than 1% of the busiest hour. The distribution of mean power with time was originally investigated by Holbrook and Dixon,[27] later by Jacobsen[28] and several telephone administrations in various parts of the world.[29] The curve obtained by Holbrook and Dixon is shown in Fig. 3. Subrizi[30] has indicated that average talker volume has apparently decreased since this original work, particularly for short-haul toll circuits. It has been pointed out that the distribution curves derived by these authors, indicate that for many design purposes an exponential relation may be assumed between the number of channels and mean power.[5] Then the following expression may be used:

$$P=XN^{\alpha} \tag{3}$$

from which

$$T_c=XN^{\alpha} \tag{4}$$

where P is the mean power, N is the number of channels, and X and α are constants. Values for X and α are given later in this section. A curve on this basis using Subrizi's data is also shown in Fig. 3. These curves do not include the effects of carrier leaks and signaling tones. The result of including spurious signals at -20 dbm 0 per channel is also given. None of the data shown include the effects of limiting. Since this tends to reduce the mean power, the curves shown may be regarded as conservative and sufficiently well approximated by the exponential relationship.

The original data were based on an observed activity factor of 0.25. This may be defined as the fraction of time a busy channel, that is, one not available for another caller, actually carries voice energy. Although this figure has been questioned,[5,31] it is reasonable to believe that this is an average figure for a typical busy circuit.

Of more significance however, is the implication that all channels are busy regardless of the number, and this is certainly sufficiently true for medium sized groups of trunks on typical exchanges. Study of traffic data,[32–34] including the latest methods of analysis by Hettwig and Rhode,[35] indicates that this is not a reasonable assumption for small numbers of channels. Fig. 4 shows trunk occupancy derived from several sources. The Rural Electrification Administration suggests a 5% grade of service for toll-connecting trunks[32] which is a reasonable qualification for scatter-circuit service. A curve on this basis is given in Fig. 4. In most cases it can be expected that the scatter channels would be routed through a manual switchboard; accordingly a curve has also been added for a 5% probability of congestion.[33] It does not seem feasible to establish a design on worse than 5% grade of service, particularly when it is recalled that the scatter link can be engineered to provide the reliability requisite of toll circuits. By contrast the CCIF (International Telephone Consultative Committee) curve for 20% deferred calls is also shown in Fig. 4, showing virtually 100% trunk occupancy.[36] It is believed that this results in 30% lost calls, which would be totally unacceptable for normal service.

By determining the number of busy channels from the average trunk occupancy indicated in Fig. 4, it is possible to derive curves for the distribution of mean power based on a specific grade of service. This has been done in Fig. 5. It is believed these curves are more representative of mean powers likely to be realized in operational systems than those of Fig. 3. Although design is ordinarily based on the curve pertaining to the 1% of the busy-hour criterion, originally this was a somewhat arbitrary choice. It represents an unnecessarily high objective

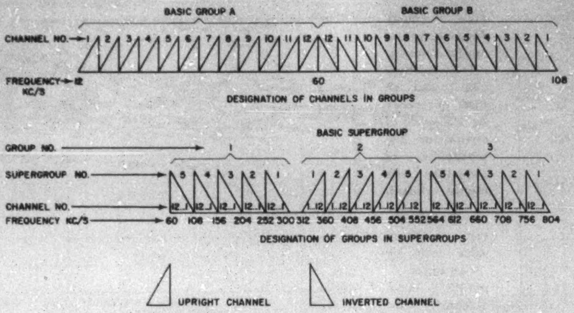


Fig. 6. CCIF channeling arrangements showing groups and super groups

for private industrial or quasi-national circuits, where somewhat reduced performance is often permissible. Accordingly, in Fig. 5, a curve for a "5% of the busy hour" criterion has been added, as a suggestion for these latter applications.

Values of the constants X and α for the various curves of Figs. 3 and 5 are shown in Table I.

Channels

In order to obtain minimum traffic capacity, it is necessary to have contiguous channels extending from $0 \rightarrow f_m$. This virtually dictates the use of CCIF frequency allocations which have tentatively been recommended for use in radio relays.[27] Channels are 4 kc wide, extending from a lowest frequency of 12 kc; see Fig. 6. Conforming equipments have been manufactured and installed in many parts of the world. When standardization, compatibility, and availability are evaluated, the use of the CCIF allocation is almost mandatory. Even when only a few channels are being considered, use of compatible equipment offers so many advantages that the value of a possible reduction in (f_m) is offset. Furthermore, the band below 12 kc is available for special purposes including service channels, so that the spectrum is normally fully utilized. Also, since this is the bottom end of the band where lower noise can be expected, it is feasible to consider as many as five additional narrow-band channels within the 12-kc limit with slightly increased transmission impairment,[29] particularly for military applications.

Table I

		X	α
Fig. 3.	A.	0.4	0.6*
	B.	0.315	0.67
	C.	0.315	0.62
Fig. 5.	p=1%	0.18	0.8*
	p=5%	0.06	0.92

*To 120 channels.

The system's analysis will, therefore, be based on the relation $Nb=f_m$ where N is the number of channels from $0 \rightarrow f_m$.

White Noise and Peak Deviation

The statistical characteristics of the complex signal comprising a number of voice channels are similar to those of white noise.[27] Although the relative peak-to-rms ratio increases as the number of channels is reduced, it can be shown that limiting substantially reduces this effect,[28] and the white-noise concept is suitable down to perhaps four to six channels. The nonlinear noise which results from the transmission of such a signal through distorting elements in the system has been analyzed for both FM and AM systems by many authors.[9,12,28,30] It can be shown that the nonlinear noise which falls in any channel is dependent on the rms value of the transmitted wave.[38] By using a white-noise signal with rms value equal to that anticipated in the working system, behavior under operative conditions is simulated. Preferably the white-noise test signal should extend over the entire base band, measurements being carried out in narrow band at the upper end of the spectrum. Data obtained with the so-called "50% illumination" are misleading, since the distribution of nonlinear noise over the base band is thereby altered and significant products may be lost.[9,38] The full base-band concept is used throughout this paper. The white-noise concept was probably originated by Bennett,[40] and has now received almost universal acceptance.[5,41] In addition it is a powerful analytic tool used by many workers in the field.[12,14,42]

It is customary to define the per-channel transmission level in terms of the peak power capability of the equipment and a loading factor.[27] This concept is of little value with the white-noise approach, since the "peak" characteristics of noise are uniquely defined by inherent statistical properties. Furthermore, it should be noted that only the rms value can be measured with any certainty, and data related to peak values can be quite misleading; this has also been observed Hamer and Medhurst.[43] This anomaly can be overcome by specifying equipment in terms of the rms noise-power handling capability, a rating which is now overdue in the industry.

It is to be noted that the CCIR has indicated a specific rms level for the white-noise testing of wide-band radio relays.[18] Recently it has been suggested that this should equal the mean power corresponding to the 1% of the busy-hour criterion as shown in Fig. 3.[44] With the test rms value given, the per-channel transmission level follows directly from the equivalent mean-power curve.

When the signal-to-noise ratio γ is measured in a channel by this particular method, a correction must be added to obtain the equivalent signal-to-noise ratio in the operative system.[5] This correction is obtained from the relation:

$$S_n=\frac{K'N\delta'\gamma}{T_c} \qquad (5)$$

where S_n is the signal-to-weighted nonlinear noise in the system and γ the signal-to-noise ratio measured in the white-noise test, and δ' has the same significance as δ in equation 1 and depends on the principal order of distortion. For the purposes of this discussion, $\delta'=2$, although it tends to be somewhat greater than this for second-order distortion[10] which is usually the most significant.

In the FM system the concept of peak deviation when modulating with a Gaussian spectrum is meaningless. In keeping with the previous discussion, the concept of rms value, and T_c, will be retained. For the purposes of calculation, the modulation index used (Δ/f_m of equation 1) will be derived from the peak deviation obtained with a sine-wave modulating signal having an rms value equal to that of the white noise. This peak deviation is therefore 1.4 times the rms deviation of the white noise. The per-channel deviation will therefore be $[(\Delta/(T_c)^{1/2}]$.

Optimum Utilization of Noise Bandwidth

Equation 1 may be written, after substituting for K', δ:

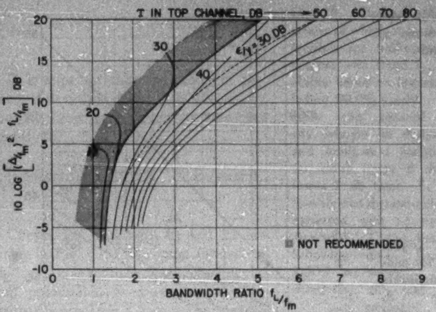


Fig. 7. The factor $[(\Delta/f_m)^2 f_l/f_m]$, in terms of bandwidth ratio and γ

$$S_t=\frac{4\epsilon f_m}{bT_c}\left[\left(\frac{\Delta}{f_m}\right)^2 f_l/f_m\right] \tag{6}$$

where $2f=B=$"truncated" bandwidth. If radian measure is used for the bracketed terms, they will be recognized as parameters in Medhurst's equations. Substituting *Nb:*

$$S_t=\frac{4\epsilon N}{T_c}\left[\left(\frac{\Delta}{f_m}\right)^2 f_l/f_m\right] \tag{7}$$

and with equations 5 and 7:

$$S_t=S_t\gamma\epsilon^{-1}\left[\left(\frac{\Delta}{f_m}\right)^2 f_l/f_m\right]^{-1} \tag{8}$$

If S_t and N are fixed, then the bracketed term of equation 7 will be fixed, so that S_n becomes a function of γ. However this is dependent on f_l/f_m. The relation between γ, f_l/f_m, and the bracketed term is shown in Fig. 7. Typical curves for S_n, derived from equations 7 and 8 and Fig. 7, are shown in Fig. 8.

A study of Fig. 7 indicates certain useful design limits. The change in direction of slope of the γ curves indicates the region in which an increase in deviation no longer produces a commensurate increase in signal-to-noise ratio; see reference 7. A smooth line has been drawn through the parameters to approximate this transition point. Values of $(\Delta/f_m)^2 f_l/f_m$ which lie above this line could not be recommended. It will be noted that this effect is less serious as γ increases, and some flexibility in the position of the nonrecommended area above $\gamma=30$ db is permissible.

A further limit is associated with the absolute value of γ. Since distortion is associated with loss of signal energy, small values of γ will be associated with noticeable level variations. Usually, small γ implies small S_t and is therefore associated with a relatively poor quality circuit, on which such variations might be acceptable. The acceptable minimum for γ is somewhat indeterminate. Taking into account the net results of this effect, the suggested minimum value of γ is 25 db, and values below 20 db could not be recommended for normal multichannel service.

In a practical design analysis, it is convenient to consider that the nonlinear noise power will be kept at some given fraction of the thermal noise power in order to examine the effect of various parameters. Accordingly,

$$S_t=yS_n \quad y<1 \tag{9}$$

It may be shown that maximum possible signal-to-total-noise ratio is obtained when the thermal and nonlinear noise powers are approximately equal (depending on the order of distortion).[8,9,13] Since nonlinear noise can be expected from several sources, the contribution of any one alone should be substantially less than the thermal noise. On the other hand, too small a value may impose unnecessarily severe restrictions on the design. It can be expected in a practical design that $0.25\geq y\geq 0.01$.

From equations 8 and 9:

$$\gamma=(\epsilon/y)\ [(\Delta/f_m)^2 f_l/f_m] \tag{10}$$

where ϵ/y is a constant for a given design decision.

The dotted curve in Fig. 7 indicates the particular relationship of equation 10 when $\epsilon/y=30$ db. Curves may be drawn in this fashion for various values of this ratio. It is found that there is a general expression for these curves, which holds very closely over the recommended area of Fig. 7. This expression is:

$$(\Delta/f_m)^2 f_l/f_m=x(f_l/f_m)^{2.5} \tag{11}$$

where x is a function of ϵ/y, and is also obtained from Fig. 7. This function is shown in Fig. 9.

The curve corresponding to the minimum permissible value of ϵ/y must not lie within the nonrecommended area of Fig. 7. This condition is satisfactorily met when $\epsilon/y=20$ db; with the previously suggested minimum for γ of 25 db, a minimum value for f_m of 5 db is obtained, yielding $f_l/f_m=1.9$. While the choice is based on a somewhat arbitrary assessment of the limits, Fig. 7 indicates that the value of 1.9 is not likely to be reduced appreciably for any reasonable re-evaluation of the limiting factors.

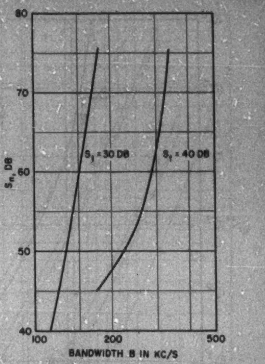


Fig. 8. S_n as function of S_t. Curves based on N=15 and curve B of Fig. 3. Effect of operating in nonrecommended region of Fig. 7 can be seen by rapid change in slope of S_t=40-db curve

When $f_l/f_m=1.9$, the minimum permissible bandwidth is obtained:

$$B=3.8\,f_m \tag{12}$$

$$B=15.2\ N\ \text{kc/second}$$

The maximum signal-to-noise ratio for this condition is obtained from equation 7 and is:

$$S_t=\frac{12.6\ \epsilon N}{T_c} \tag{13}$$

If full advantage is taken of compandors, it should be possible to reduce the ordinate of Fig. 7 to zero, yielding $(f_l/f_m)\simeq 1.5$ and giving $B=3\,f_m$. Although this is only a small reduction in bandwidth, the improvement in signal-to-noise ratio would be significant.

Values used for ϵ vary from 9 to 12 db among different authors.[45,46] As indicated by Crosby's work[47] it depends on the relative position of the channel on the base band, being greatest for the lowest channel, and Harris[48] has given a precise relation on this basis. The value determined for various multichannel systems varies very little and 10 db for the top channel is a suitable figure for design and analytical purposes. Using this, and equation 4 in equation 12:

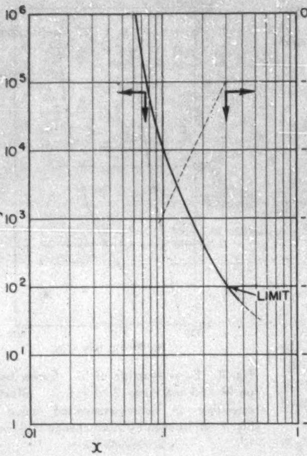


Fig. 9 (left). Relations between x, y, and ε

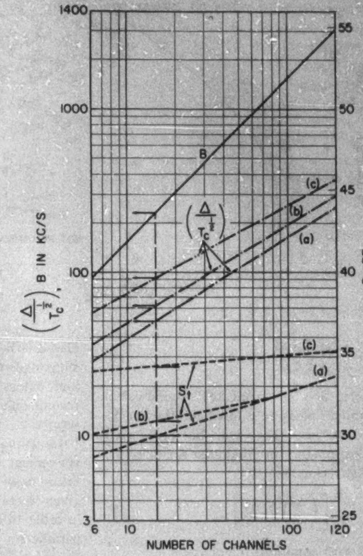


Fig. 10 (right). Relations for minimum permissible bandwidth B, which, in practical design, should be increased by at least 10% over value obtained from these curves

a—Based on curve B, Fig. 3

b—Based on Fig. 5, p=1%

c—Based on Fig. 5, p=5%

$(\Delta/T_c^{1/2})$ = per-channel deviation

$$S_t = \frac{126 N^{1-\alpha}}{X} \tag{14}$$

This indicates that compandors would be needed on minimal bandwidth systems if toll-quality circuits were required. Since ϵ/y is fixed at 20 db, then $y=0.1$ for all values of S_t. This is quite satisfactory and in keeping with stipulated limits. Also, since $f_l/f_m=1.9$, then $\Delta/f_m=1.29$.

Thus, the per-channel deviation, for maximum signal-to-noise ratio is:

$$(\Delta/T_c^{1/2}) = \frac{5.16\, N^{(1-0.5\alpha)}}{X^{0.5}} \text{ kc/second} \tag{15}$$

Equations 12, 14, and 15 are shown in Fig. 10.

It is possible to derive more general relationships for B in terms of S_t. Using equations 7 and 11:

$$S_t = \frac{4\epsilon N x}{T_c}(f_l/f_m)^{3.5} \tag{16}$$

To a close approximation, from Fig. 9,

$$y = 1.25\, x^4 \tag{17}$$

It is to be noted that if a corresponding equation is developed for the range in which $y<0.01$, even higher orders of x are indicated.

From equations 16 and 17:

$$S_t = \frac{3.78\, \epsilon^{0.75} N y^{1/4}}{T_c}(f_l/f_m)^{3.5} \tag{18}$$

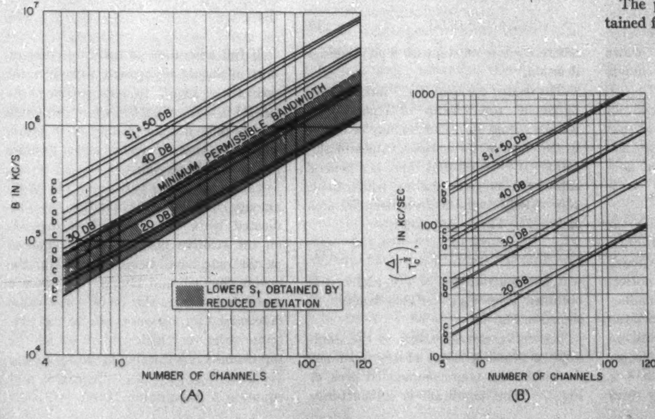
Using $\epsilon=10$ and substituting for f_l:

$$B/f_m = \frac{0.83}{y^{1/4}}\left(\frac{S_t T_c}{N}\right)^{1/3.5} \tag{19}$$

Substituting for f_m and using the values in Table I:

$$B = \frac{3.32\,(XS_t)^{1/3.5} N^{\left(\frac{\alpha+2.5}{3.5}\right)}}{y^{1/14}} \text{ kc/second} \tag{20}$$

The per-channel deviation can be obtained from equations 6 and 20 giving:



(A)

(B)

Fig. 11. Curves for equations 20 and 22. All curves based on y = 0.1

A—Relation between optimum bandwidth and signal-to-noise ratio B in kc per second

a—based on curve B, Fig. 3

b—based on Fig. 5, p=1%

c—based on Fig. 5, p=5%

B—Per-channel deviation for A

$$\frac{\Delta}{T_e^{1/2}} = 0.98(S_t^{3.5} y^{1/4} N(4.5-\alpha) X^{-1})^{1/7} \text{ kc/second} \quad (21)$$

It is apparent that the choice of per-channel deviation is but little affected by changes in y so equation 21 simplifies slightly to:

$$\frac{\Delta}{T_e^{1/2}} = 0.88(S_t^{3.5} N^{(4.5-\alpha)} X^{-1})^{1/7} \text{ kc/second} \quad (22)$$

Curves for equations 20 and 22 are given in Fig. 11. The most interesting observation is that the per-channel deviation permissible is much greater than ordinarily used, while the noise bandwidth B is narrower. This implies that much lower transmitter power is needed than presently applies. These equations and their method of derivation also indicate that by controlling bandwidth at the transmitting terminal, side-band "spread" might be reduced with consequent improvement in spectrum conservation. In view of the substantial dependence of y on B, it is suggested that minimal practical bandwidths should be at least 10% greater than obtained from equation 20 in order to provide a sensible design margin.

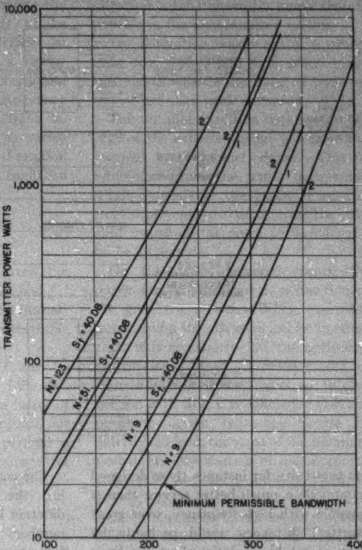


Fig. 12. Optimal transmitter power of which at least half is feasible for traffic-controlled systems with fixed f_m. Reliability based on instantaneous fading, 99.9% yearly. Quadruple ratio squared diversity. Margin for design error and equipment variations 8 db, noise figure 8 db, line losses 5 db total. S_t signal-to-weighted thermal-noise ratio expected with r-f signal at threshold. Curves based on following data: frequency 1,000 mc/second, 60-foot parabolas, 60 feet above ground

1—Traffic expected for 95% of busy hour with 5% grade of service

2—Traffic expected for 99% of busy hour using curve B, Fig. 3, full trunk occupancy

Power and Reliability

It is convenient to express the required transmitter power W_t in the form:[1]

$$W_t A = P_t \quad (23)$$

where A is a factor determined by the total r-f transmission loss from transmitter output to receiver input for a given reliability and antenna configuration. The determination of A is beyond the scope of this paper but is readily obtained from established procedures.[49] Using equations 2 and 12 with $\epsilon = 10$:

$$W_t = \frac{F}{A} N \; 10^{-18.2} \text{ kw} \quad (24)$$

The general case is obtained from equation 20. It is permissible to use an average value for y since its effect is so small. Then:

$$W_t = \frac{F\,10^{-17.58}}{A}(XS_t)^{1/3.5} N^{\left(\frac{\alpha+2.5}{3.5}\right)} \text{ kw} \quad (25)$$

By choosing definite equipment configuration, A can be expressed in terms of distance over a smooth spherical earth, and the required power can then be expressed for various performance criteria. This has been shown in Fig. 12.

It has already been suggested that there is need for a 100-watt system.[1] This is substantiated by Fig. 12, and it would appear that 100-watt and 250-watt equipment should be added to the present range. It should be possible to prove in systems in this category in many parts of the world at present in need of reliable circuits.

Equation 25 provides an interesting index of system efficiency. Since W_t is a figure derived for an optimized system efficiency can be expressed in terms of W_t by the relation:

$$E_f = \frac{W_t}{\text{Transmitter power called for}}$$

The absolute value for S_t is still open to considerable discussion,[50] although acceptable values for commercial toll service have been discussed.[51] Recently it has been suggested that for the toll service, S_t should be greater than 40 db for 99% of the busy hour. A value of 35 db for 95% (radio equipment alone) is suggested for services such as private industry and utilities.[44] The curves of Fig. 12 show these two classes of service, on the basis of trunk occupancy established for 5% grade of service, as shown in Fig. 5.

The question of the figure which should be used for propagation reliability, in the design of scatter circuits, is not yet resolved, though 99.9% seems to be fairly well accepted. A considerable amount of information has been published relating to the so-called long-term and short-term effects.[52] A study of fading rates[60] indicates that, in general, the short-term effect is slower than the integration time of the ear, so that one would expect the subjective effect to be determined by the instantaneous level of noise. The instantaneous fading distribution requires the solution of a bivariate probability integral, and curves on this basis have been derived.[4] Investigation of these data leaves little doubt as to the upper design limit. The slope of these distributions increases sharply about 99.95% reliability and implies that a disproportionate increase in r-f power is required for better performance. A practical system-design limit then, is 99.9% which corresponds to an "outage" of about 8 hours per year. Fig. 12, which is based on Norton's data[49] makes use of this limit. It is doubtful if much is to be gained by reducing reliability below this figure when the period involved is taken into account. Due to the long cycle in the fading phenomena, the reliability is expressed on an annual basis. A change from 99.9% to 99% for instance, over an average path length and with a quadruple diversity system, represents a possible decrease of no more than about 6 db in power for a loss of 3 days outage, a figure which would be quite unacceptable for good quality circuits. However, low-

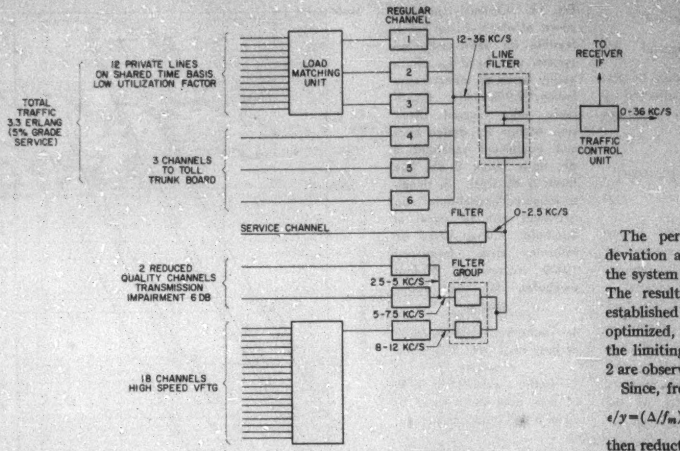

REGULAR CHANNEL
12 PRIVATE LINES ON SHARED TIME BASIS LOW UTILIZATION FACTOR
TOTAL TRAFFIC 3.3 ERLANG (5% GRADE SERVICE)
LOAD MATCHING UNIT
1
2
3
4
5
6
12-36 KC/S
LINE FILTER
TO RECEIVER IF
0-36 KC/S
TRAFFIC CONTROL UNIT
3 CHANNELS TO TOLL TRUNK BOARD
FILTER
0-2.5 KC/S
SERVICE CHANNEL
2 REDUCED QUALITY CHANNELS TRANSMISSION IMPAIRMENT 6 DB
FILTER GROUP
2.5-5 KC/S
5-7.5 KC/S
8-12 KC/S
18 CHANNELS HIGH SPEED VFTG

Fig. 13. One manner of fully utilizing small base band. N=9. Very high commercial utilization of scatter medium obtained with this arrangement

grade circuits, for instance those designed to provide a much better service than is possible with high frequency, over great (scatter) distances would probably be very acceptable if designed on the basis of 98%; i.e., about 1 week total outage time per year.

It is to be borne in mind that all these figures refer only to the disappearance of signal below threshold. Most circuits will "hold" down to about $\epsilon=2$, and since the design is predicted on $\epsilon=10$, the actual "lost" circuit time, even for systems designed for 98%, is likely to be less than 4 hours total.

Economic Index

Economic evaluation of the scatter circuit requires some knowledge of the price which must be paid for a given performance criterion. An indication both of capital cost and technical and logistic difficulty is obtained from a knowledge of the required power. One method which appears to be worth following is to use equations 25 and 4. It is to be recalled that W_t in equation 25 represents the minimum permissible power for specified system parameters and equipment configurations. Then:

$$\frac{(T_e\, S_t)^{1/1.8}}{N\,A}=W_t\left(\frac{y^{1/16}}{F\,10^{-15.9}}\right) \qquad (26)$$

The factors T_e, S_t, N, and A, are all related to performance criteria, while the bracketed quantity on the right-hand side is a constant. Thus the left-hand side is a measure of the technical and economic feasibility and may be referred to as an "economic index."

The concept is quite useful in comparing the proportionate economics of different links in a network when various numbers of channels and grades of performance are involved.

In addition, since A is determined by propagation reliability, the effect of changing the limit on this factor can be assessed.

Utilization of Optimal Condition

Certain techniques for the reduction of the power required for scatter circuits can be examined with the aid of the previous text. One manner of using fully a small base band is shown in Fig. 13.

Controlled Carrier Operation

A method of adjusting transmitter power by means of a pilot signal over a reciprocal path has been described.[4] The equations developed in the present paper indicate only that possibly the maximum power required for a given service may be less than that presently used. There is no indication that improvement in this technique can be effected by the concepts developed here.

Control by Noise

In equations such as 20, S_t can be replaced with $S_t'r$, where S_t' is the signal-to-noise ratio obtained when fading is taken into account, and r is the ratio of the received r-f signal during fading to threshold signal. When r is $>>1$, then $S_t'>>S_t$, and there is the possibility that the bandwidth might be reduced slightly so that the total noise (thermal plus nonlinear) remains constant.

The per-channel and resultant rms deviation are based on specific values of the system parameters, as in equation 6. The resultant nonlinear noise is then established from Fig. 7. If the design is optimized, then, as already indicated, the limiting conditions indicated by Fig. 2 are observed.

Since, from equation 10:

$$\epsilon/y=(\Delta/f_m)^2\,(f_l/f_m)$$

then reduction of bandwidth (f_l) must be accomplished by an increase in y. Since the optimized design is already based on maximum tolerable y, the method is open to serious question. Even more serious is the fact that the nonlinear noise, in the range of interest, varies with the fourteenth power of bandwidth, which can be seen both from equation 20 and from Fig. 1. Consequently, either the available control must be too small to be of use, or else y must be much too small in the initial design. It would seem, therefore, that this technique does not have anything to offer the optimized system, though it may be useful in "over-engineered" designs.

Control by R-F Signal

Since output thermal noise decreases with strong r-f signals, there is the possibility of using the received carrier to reduce bandwidth during these periods. However, the same remarks apply as for the previous paragraphs. In this case, since there is no correlation between r-f signal and traffic, any method of control is fundamentally unsound. Variations in r-f signal are most effectively utilized by means of the controlled carrier techniques mentioned earlier.

Control by Traffic

The long-term mean power of the multichannel signal is dependent on the number of simultaneously busy channels and, therefore, on traffic load. This exhibits a considerable diurnal variation[44] which may be used to advantage.

Fixed f_m

The design value of S_t is determined by affixing suitable values to the various factors of equation 6. In practice of course, Δ is a variable, determined by actual traffic. Let Δ' represent this variable in the operative system.

Equation 11 may then be rewritten in the following way:

$$\left(\frac{\Delta'}{f_m}\right)^2=(B/2f_m)^{2.5} \tag{27}$$

The variation expected in $(\Delta')^2$ can be obtained from mean power distribution curves of which Figs. 3 and 5 are an example. This variation depends on the number of channels and the residual level for spurious tones. It is quite large for systems of less than 60 channels designed on a 1% of the busy-hour basis. A "ball park" figure is 10 db, which permits B to be reduced by the factor 2.5. The net power saving is obtained by plotting the power transmitted as a function of the diurnal variation of B, based on the variation of (Δ'). Performing this for a typical system indicates that total daily transmitted power requirement is reduced to about one half of the uncontrolled value.

Of more interest, perhaps, is the improvement in reliability which can be obtained. Diurnal fading is small, so that reduction of outage obtained by reducing the bandwidth represents a commensurate improvement in reliability. Reference to traffic volumes[44] indicates that full bandwidth would be required for less than 12 hours. On this basis, the system could be designed for approximately twice the failure rate, i.e., 99.8%, with expectation that performance would be equal to that of a 99.9% uncontrolled system. This represents a permissible reduction of about 3 db in actual transmitter power for typical scatter angles.

Devices for this particular function are not yet available. They should have a fast (5 minutes) time constant to open the bandwidth and a slow one (at least 1 hour) to close it. It is of interest to observe that, if units are used at transmitter and receiver as suggested above, synchronism is not necessary.

Reduction in spurious signal (baseband) power (i.e., carrier leaks) does not achieve a great deal. The minimum permissible value for f_t/f_m is probably 1.5, when Δ' is small. With an optimized system, and for typical values of S_t, the range of bandwidth variation permissible is relatively small. Where a large range appears to be feasible, the implication is that the system is not optimized in the first place.

Controlled Channel Selection

The previous study does not take advantage of reduced trunk occupancy as traffic flow decreases. This permits the number of available channels to be reduced in accordance with traffic demand. In turn this implies a device which selects the number and position of channels, so that the lower frequency channels are utilized when the traffic demand is small. Therefore, f_m is varied in accordance with traffic. Equation 27 can be rewritten for this condition giving:

$$B=\left[\frac{2(\Delta')^2(f_m')^{1.5}}{x}\right]^{1/2.5} \tag{28}$$

As before, Δ' is determined from the diurnal variation of mean power and f_m' is a variable obtained from the (diurnal) circuit demand in accordance with traffic flow.

This concept indicates that B may be reduced even further than indicated under "Fixed f_m." A reduction factor of 10–20, depending on design, appears not unreasonable. This in turn permits a still greater reduction in transmitter power or an effective increase in propagation reliability.

LOAD MATCHING

The possibility of using units which provide a function similar to that of line concentrators has been mentioned.[1] It is possible to consider such a device in which specific entrance circuits are always routed to specific output circuits. This could increase effective channel capacity provided the minimum traffic flow did not exceed the limit for the "concentrating" circuits. This may be thought of as matching the load offered by the connected lines to the permitted capacity of the channels. Increased private-line facility would be possible, and this may have considerable advantage in many instances where there is a limited channel capacity and circuit rental is feasible. This technique also lends itself to a minimum first-cost system, with expansion predicted by demonstrated utilization. Although this concept does not seem to have yet been exploited, it is in no way restricted by the procedures of the present paper.

Conclusions

The nonlinear noise introduced by truncating the bandwidth may be used as a limiting factor in the analysis of the optimized multichannel system. The signal-to-thermal-noise ratio may then be expressed as a function of traffic volume and noise bandwidth. This leads to a minimal bandwidth criterion which permits substantially greater signal-to-noise ratio and lower transmitter power than is possible in the nonoptimized system.

The derivation makes use of a white-noise test signal, and it is pointed out that this should be expressed in terms of rms rather than peak deviation.

The optimal condition requires knowledge of the expected mean power of the complex multichannel signal. This is established from the trunk occupancy, which depends on the grade of service, and the percentage of time during the busy hour for which the given performance is to be expected.

On this basis, required transmitter power can be expressed as a function of the grade of service. When expressed in terms of a smooth spherical earth reference, it is concluded that the optimized system lends itself to development of 100- and 250-watt installations. In connection with this, it is concluded that the optimum design reliability should be determined from the slope of the bivariate probability distribution of instantaneous fading, leading to a limiting value of 99.9%, expressed on a yearly basis. This concept also leads to the use of a system efficiency factor which may be expressed in terms of "optimized" transmitter power. In addition, an economic index can be established which gives a measure of the economic and technical feasibility of the system in terms of the principal performance criteria. Various methods of reducing transmitter power by establishing variable parameters are received on the basis of the optimal design. A new technique is suggested in which traffic volume controls the noise bandwidth, and it is believed that the transmitter power can be reduced substantially by this procedure.

One general conclusion may be drawn from a study of the data presented here: it would appear that an unnecessarily high price still must be paid for the nonavailability of equipment tailored to the optimal requirements of the system. On the other hand, the optimized system, together with the capabilities of traffic controlled bandwidth and the use of load-matching devices to increase potential revenue, should permit commercial breakthrough for the tropospheric-scatter technique.

References

1. Factors Affecting the Use of Over-the-Horizon Links in Telecommunication Networks, Charles A. Parry. *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 77, Sept. 1958, pp. 485–96.

2. Interests and Inflation in the Economics of Comprehensive Comparisons in the Planning of Telecommunication Systems, R. Krzycz-

kowski. *ATE Journal*, Liverpool, England, Oct. 1957, p. 298.

3. Economic Analysis in Long Term Planning of Military Communication Systems, **R. Krzyczkowski.** *WESCON Convention Record*, Institute of Radio Engineers, New York, N. Y., pt. 8, 1958.

4. Basic Analysis on Controlled Carrier Operation of Tropospheric Scatter Communication Systems, **L. P. Yeh.** *Convention Record*, Institute of Radio Engineers, pt. 8, p. 261.

5. Design Factors for the Optimization of Multichannel Radio Systems, **Charles A. Parry.** *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 76, Nov. 1957, pp. 606–20.

6. A Mathmatical Theory of Communication, **C. Shannon.** *Bell System Technical Journal*, New York, N. Y., July 1958, p. 379.

7. Theoretical Considerations on the Rate of Transmission of Information, **W. Tuller.** *Proceedings*, Institute of Radio Engineers, May 1959, p. 468.

8. Communications in the Presence of Noise, **C. Shannon.** *Ibid.*, Jan. 1949, p. 10.

9. Microwave Radio Links, **A. T. Starr, T. H. Walker.** *Proceedings*, Institution of Electrical Engineers, London, England, pt. III, no. 61, Sept. 1952, p. 241.

10. The Equalization of Baseband Noise in Multichannel FM Radio Systems, **C. A. Parry.** *Proceedings*, Institute of Radio Engineers, Nov. 1957, p. 1527.

11. The Application of Formulas for Distortion of Frequency Modulation to VHF Multichannel Radio Telephony, **J. Merriman, R. White.** *Proceedings*, Institution of Electrical Engineers, pt. IIIA, 1947, p. 649.

12. Interchannel Interference in FM and PM Systems Under Noise Loading Conditions, **W. R. Bennett, H. E. Curtis, S. O. Rice.** *Bell System Technical Journal*, May 1955, p. 601.

13. VIII Plenary Assembly, CCIR, Warsaw, Poland, vol. I, 1956.

14. RF Bandwidth of Frequency Division Multiplex Systems Using Frequency Modulation, **R. G. Medhurst.** *Proceedings*, Institute of Radio Engineers, New York, N. Y., Feb. 1956, p. 189.

15. Normalized Phase and Gain Derivations as an Aid in Evaluation of FM Distortion, **J. J. Hupert.** *Ibid.*, p. 438.

16. Frequency Modulation Distortion in Linear Networks, **R. F. Brown.** *Proceedings*, Institution of Electrical Engineers, pt. B, Jan. 1957, p. 52.

17. The Distortion of FM Signals in Passive Networks, **R. H. Collings, J. K. Swirzynaski.** *Marconi Review*, St. Albans, Hertfordshire, England, vol. 17, 1954, p. 113.

18. The Transmission of a Frequency Modulated Wave Through a Network, **W. J. Franz.** *Proceedings*, Institute of Radio Engineers, Mar. 1956, p. 114.

19. A Theoretical and Experimental Investigation of Tuned Circuit Distortion in Frequency Modulation Systems, **D. Jaffe.** *Ibid.*, May 1945, p. 171.

20. Distortion of FM Waves by Transmission Networks, **A. S. Gladwin.** *Ibid.*, Dec. 1947, p. 1436.

21. Harmonic Distortion of Frequency Modulated Waves by Linear Networks, **R. G. Medhurst.** *Proceedings*, Institution of Electrical Engineers, pt. III, May 1954, p. 171.

22. Distortion of Frequency Modulated Signals in Electrical Networks, **F. L. Stumpers.** *Communication News*, Camden, N. J., vol. 9, 1948, p. 82.

23. *Discussion* of reference 20 by **F. L. Stumpers, A. S. Gladwin.** *Proceedings*, Institute of Radio Engineers, 1948, p. 1257.

24. Distortion Band Pass Considerations in Angular Modulation, **A. A. Gerlach.** *Ibid.*, Oct. 1950, p. 1203.

25. Universal Design Circuits for Intermediate Frequency Amplifiers with Particular Reference to Phase and Amplitude Distortion and Their Compensation, **W. Hackett.** *Ibid.*, Dec. 1950, p. 1408.

26. Vacuum Tube Amplifiers (book), **George E. Valley, Jr., Henry Wallman,** editors. McGraw-Hill Book Company, Inc., New York, N. Y., 1948.

27. Load Rating Theory for Multichannel Amplifiers, **B. D. Holbrook, J. T. Dixon.** *Bell System Technical Journal*, vol. 18, 1939, p. 624.

28. The Effect of Nonlinear Distortion in Transmission Systems, **B. B. Jacobsen.** *Electrical Communication*, New York, N. Y., vol. 19, 1940, p. 29.

29. XVII Plenary Assembly, CCIF, Geneva, Switzerland, vol. I, 1954.

30. A Speech Volume Survey on Telephone Message Circuits, **V. Subrizi.** *Bell Laboratories Record*, New York, N. Y., Aug. 1953, p. 292.

31. Single Sideband Techniques in VHF Long Range Communications, **W. E. Morrow, C. L. Mack, B. E. Nichols, J. Leonhard.** *Proceedings*, Institute of Radio Engineers, Dec. 1956, p. 1854.

32. Telephone Engineering and Construction Manual. Rural Electrification Administration, Washington, D. C., issue 4, sect. 510, Apr. 1956.

33. Traffic Calculations for Trunk and Local Telephone Exchanges. *Pamphlet 543A*, Siemens Brothers and Company, Ltd., London, England.

34. Theories for Toll Traffic Engineering in the USA, **R. I. Wilkenson.** *Bell System Technical Journal*, Mar. 1956, p. 421.

35. New Reference Data for Calculating Trunking and Switch Quantities for Automatic Telephone Exchanges, **E. Hettwig, K. Rhode.** *Siemens-Zeitschrift*, Erlangen, Germany, vol. 30, Jan. 1956, p. 1.

36. XVII Plenary Assembly, CCIF, Geneva, Switzerland, vol. VI, Oct. 1954.

37. *Ibid.*, Vol. III.

38. Non-Linear Distortion in Transmission Systems, **R. A. Brockbank, G. A. Wass.** *Proceedings*, Institution of Electrical Engineers, pt III, Mar. 1945, p. 1.

39. Distortion in Frequency Modulation Systems Due to Small Sinusoidal Variations of Transmission Characteristics, **R. G. Medhurst, G. F. Small.** *Proceedings*, Institute of Radio Engineers, Nov. 1956, p. 1608.

40. Cross Modulation Requirements on Multichannel Amplifiers Below Overload, **W. Bennett.** *Bell System Technical Journal*, Oct. 1940, p. 587.

41. The White Noise Test Method of Measuring Crosstalk and Interference in Multichannel Telephone Link Systems, **J. F. Golding.** *Electronic Engineering*, London, England, May 1958, pp. 3, 9.

42. an Extended Analysis of Echo Distortion on the FM Transmission of Frequency Division Multiplex, **R. G. Medhurst, G. F. Small.** *Proceedings*, Institute of Radio Engineers, pt. B, Mar. 1956, p. 190.

43. Correspondence, **R. Hamer, R. G. Medhurst.** *Ibid.*, Dec. 1956, p. 1878.

44. CCITT Recommendations for Multichannel Radio Relays and White Noise, **Charles A. Parry.** *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 78, May 1959, pp. 107–17.

45. System Parameters Using Tropospheric Scatter Propagation, **H. H. Bevarage, E. A. LaPorte, L. C. Simpson.** *RCA Review*, Princeton, N. J., vol. 16, no. 3, Sept. 1956.

46. Radio Engineering Handbook (book), **K. Henney.** McGraw-Hill Book Company, Inc., fourth edition, 1950.

47. Frequency Modulation Noise Characteristics, **M. G. Crosby.** *Proceedings*, Institute of Radio Engineers, Apr. 1937, p. 472.

48. An Expanded Theory for Signal-to-Noise Performance of FM Systems Carrying Frequency Division Multiplex, **D. P. Harris.** *Convention Record*, Institute of Radio Engineers, pt. 8, 1958, p. 298.

49. The Use of Angular Distance in Estimating Transmission Loss and Fading Range for Propagation Through a Turbulent Atmosphere Over Irregular Terrain, **K. A. Norton, P. L. Rice, L. E. Vogler.** *Proceedings*, Institute of Radio Engineers, New York, N. Y., Oct. 1955, p. 1488.

50. Controlled Carried Communications. *Final Report*, Electronics Division, Westinghouse Electric Corporation, Baltimore, Md., Contract no. AF19(604)-2194.

51. Design of FM Radio Relay Systems for Multichannel Telephone Service, **J. W. Halina.** *Transactions*, Professional Group on Communications Systems, Institute of Radio Engineers, Oct. 1956, p. 3.

52. *Proceedings*, Institute of Radio Engineers, "Scatter Propagation Issue," Oct. 1955.

53. Prediction of the Cumulative Distribution with Time of Ground Wave and Tropospheric Wave Transmission Loss, **P. L. Rice, A. G. Langley, K. A. Norton.** *Report*, National Bureau of Standards, Washington, D. C., June 1958.

Tubes or Transistors: A Realistic Assessment

R. E. MOE
NONMEMBER AIEE

THE POPULAR NOTION that transistors will undoubtedly replace vacuum tubes in the near future has been widely publicized partly because of the natural enthusiasm and optimism of the many new firms starting into the business of designing and manufacturing semiconductor devices. There is, of course, no doubt that transistors will take over some of the application areas previously served by tubes, and that there will be some areas where either tubes or transistors will serve equally well, but there are many fields where tubes are and will continue to be superior. The advantages and disadvantages of each device, under the wide variety of operating conditions and circuit requirements, should govern the choice for a particular application, and the following discussion is intended to bring out those which are pertinent to such an analysis.

The purpose of this discussion is to present briefly a few extracts from the available data, and list references giving more complete information, to allow a realistic assessment of performance and reliability factors of interest to the designer of electronic equipment. Broadly, these break down into two groups: entertainment, on the one hand, including radio and TV (television), and industrial and military on the other. In the first case, cost will continue to be the governing factor dictating the use of tubes except in the personal portable field, and the output stages of car radios. In the industrial and military fields, where cost is less of a factor, the relative performance and reliability factors must be examined critically, in order to resolve this tendency to play up and perhaps oversell the glamour of this newcomer to the field of electronic devices.

Accordingly, the following points have been chosen as prime factors in this analysis to guide the designer's choice, depending, of course, on his equipment specification requirements:

1. Power input (efficiency)
2. Ambient or hot spot temperature
3. Upper frequency limit
4. Noise figure
5. High-voltage requirement
6. High power output requirement
7. Spread of characteristics and tolerances
8. Combination and multiple units
9. Nuclear radiation
10. Reliability
11. Physical size and weight

The following paragraphs give the major pros and cons for each of these points, and a few pertinent quotations and references. More detailed information is available from the references.

Power Requirements

If available power input is limited, and high power output is not a prime requirement, transistors are the logical choice. For example, in hearing aids, personal portable radios, or even space satellites, which are supplied by battery power, the low drain and high efficiency of these devices are the prime factors. On the other hand, if the equipment is to be a-c operated, these are of little consequence, and other factors will govern; these may well favor tubes over transistors. For example, for fixed station receivers and transmitters, or multiple PA (public address) systems, tubes should be simpler and more economical.

Ambient Temperature

Because of the inherent nature of semiconductors, their characteristics are far more sensitive to changes in temperature than are tubes, so that germanium and to a lesser extent silicon units show considerable variation in parameters with changes in temperature, and reach a limiting capability around 100 C (degrees centigrade) and 200 C junction temperature, respectively. Tubes, on the other hand, can operate well above 200 C bulb temperature in the glass types, and over 400 C in ceramic, with practically no change in characteristics. Thus, the use of transistors under varying ambient conditions requires a great deal of attention to temperature compensation elements and feedback circuits where in most cases these precautions are unnecessary when using tubes.[1] In fact, the serious degradation in performance of transistors with increasing temperature may well make their use uneconomical because of the low gain per stage, or limited output available at the high end of the operating temperature range, well before the advertised temperature limit is reached.

Another illustration is the necessity for precautionary measures when soldering transistor leads, to prevent overheating the junction and damaging it permanently.

To quote Captain W. I. Bull: "Presently the temperature requirements for components range from −65 C to +200 C. Guided missile and Mach 2-plus manned aircraft may require components capable of providing reliable operation in ambients of 250 C to over 500 C. The goal for the development of an entire line of high-temperature, 500 C component parts is 1965. In view of the temperature limitations of germanium and silicon transistors, new semiconductor materials will have to be investigated [and this means developed, perfected, and reduced to practical manufacture—R.E.M.], and device development programs must be initiated on the most promising materials if the military's high-temperature part objective is to be met. Of course, ceramic electron tubes, to mention a dirty word, will be available in the event high-temperature transistors have not been developed."[2]

Frequency

At the present time, and with presently available constructions and techniques, transistors become increasingly difficult and expensive to build for operation above 500 kc. Refinements, of course, are being introduced, but there is a fundamental conflict between barrier thickness, area, and carrier mobility, versus capacitance, gain, and power handling ability (heat dissipation) with increasing frequency. Tubes, of course, have these problems to some extent owing to transit-time effects and interelectrode capacitance loading, but to a much smaller extent. Even the most optimistic predictions by transistor manufacturers provide little hope that they will take over present vhf (very-high-frequency) and uhf (ultrahigh-frequency) tube applications, from 100 to 1,000 mc (megacycles), in the forseeable future, unless other considerations than cost and gain govern the choice. As an example, some mobile communications equipment are now being transistorized. There are

---

Paper 60-73, recommended by the AIEE Electronics Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., January 31–February 5, 1960. Manuscript submitted June 5, 1959; made available for printing December 8, 1959.

R. E. Moe is with the General Electric Company, Owensboro, Ky.

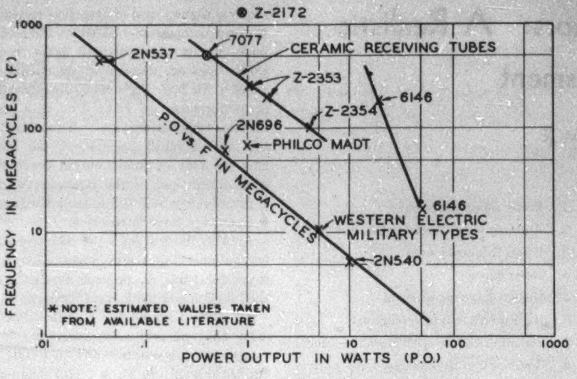


Fig. 1. Tube and transistor output versus frequency

very few transistor types registered for this kind of operation, and none with approved military specifications; see Fig. 1.

Also, several set manufacturers have produced experimental samples of transistorized portable TV receivers, but they will admit that it was necessary to exercise a great deal of selection to find the right kind of transistor to perform several of the more critical functions; this would certainly not be practical at this time.

Noise

Although relatively free of microphonics and variable leakage path noise contributions, transistors have an inherently higher random noise output than tubes, when considered on a noise power basis ($KT\Delta f$). Thus transistors may well be chosen for low level stages of audio preamplifiers to eliminate microphonics and hum, but be unsuitable for wide-band i-f (intermediate-frequency) or r-f (radio-frequency) amplifiers in radar or television, where a noise figure approaching the theoretical minimum is a prime requirement. The ceramic *7077* is appreciably better than the best transistor for ultra-low-frequency flicker noise; see Fig. 2.

High Voltage

Because of the inherently better insulation afforded by vacuum devices, it is expected that tubes for high-voltage applications, both diodes and multi-element, will be simpler, more economical, and more reliable than semiconductors for some time to come. Such things as high-voltage sweep tubes, rectifiers, and dampers for supplying 20 kv to television picture tubes or radar indicator scopes will be difficult to displace, unless some more efficient methods are found which can tolerate the lower breakdown voltage limitations of semiconductors.

Another aspect of this same weakness is found in the inability of transistors to withstand power supply transients as experienced in aircraft generator systems, which have been found to vary between −25 and +80 volts on a 28-volt d-c system,[3] or diode breakdown when using a continuity meter to check out a circuit, inadvertently applying excess reverse voltage to the semiconducting element. In fact, the elaborate protective devices necessary in the former case make such applications of dubious value.

To quote N. W. Caquelin: "In summary it can be said that transistors and other semiconductor devices can be successfully used in environments exhibiting the hazard of extreme positive and negative voltage transients *if* proper design techniques are used. It must be recognized that protection of equipment from transients, especially that using power transistors, represents some sacrifice in required space (size or parts density), efficiency (power dissipation and box temperature), equipment cost (cost of added components and cooling system), and to a lesser degree, weight. As the state of the art in the use of transistors is still relatively young, more desirable protection systems or higher voltage transistors may be developed in the future. Until then, we must either eliminate the transient from the primary power source or else accept the additional components of the protective systems. The reduction of primary power transients would undoubtedly improve the reliability of other components in airborne equipments."[3]

High Power

Because of the elaborate measures necessary to cool the junction of a power transistor, and the relatively expensive components necessary to filter and regulate a low-voltage high-current supply, vacuum tube power amplifiers will probably continue to have an economic advantage. A modifying factor here, however, is introduced by the possibility of eliminating the output transformer in the case of low-impedance loads, such as the voice coil of a loud-speaker, allowing this saving to balance out the difference noted above. The choice then rests on a companion characteristic, linearity, which is easier to achieve with tubes than with transistors.

In addition to the fact that linear characteristics are easier to achieve with tubes, transistors have the added difficulty of variable input and output impedance from small to large signals, making driving and matching difficult, and adding to the distortion.

Spread and Control of Characteristics

At present, the properties of semiconductors are more difficult to control in production than those of tubes, resulting in a much wider spread of product characteristics. In most cases, this spread is too wide for any one intended application, requiring the selection and labeling of several different portions of the over-all product falling into various groupings for gain, breakdown voltage, etc. This has produced a tremendous multiplicity of type numbers (over 700, not counting experimental), with various combinations of these properties, and, depending on the yield in each area, widely varying costs. Some of this variation will undoubtedly be reduced by further manufacturing refinements, but some of it may be inherent in the nature of the device, because of the sensitivity of the semiconducting material to extremely minute amounts of contamination. As a result, where spread of characteristics or close tolerances, or the ability to turn out a uniform product over an extended period is a prime consideration, tubes should continue to be superior.

For example, the custom is to release a quantity of type numbers for transistors made with a given physical arrangement, and test them to identical limits and characteristics, except for one critical parameter. In other words, it is impossible to control this latter characteristic, so it must be accomplished by selection. For instance, the following groups of types are identical except for alpha cutoff frequency!

2N444—0.5 mc
2N445—2.0 mc
2N446—5.0 mc
2N447—9.0 mc
2N519—0.5 mc
2N520—3.0 mc
2N521—8.0 mc
2N522—15.0 mc
2N523—21.0 mc

This is obviously a serious obstacle to standardization.

For further evidence, the following news item is of interest: "Transistor Preferred Circuits. National Bureau of Standards program to develop preferred transistor circuits, similar to preferred tube circuits developed earlier, has run into a snag because of extremely wide variability of characteristics in transistors. Whereas transconductance of a group of Mil-standard tubes normally varies only ±10%, corresponding important characteristics of transistors, such as short-circuit current gain, may vary 200% or more."[4]

Also, examination of the few military specifications available for transistors will reveal that they control only a very few characteristics, and these to very much looser AQL (Acceptable Quality Level) values than the corresponding tube characteristics. By contrast, military specifications for tubes now represent an extremely good guarantee of reliability over a 1,000-hour period.

The really basic consideration here is admirably outlined and explained in reference 5. While the entire article is well worth reading, the main point is that in intrinsic devices, the various factors such as frequency, power density, coherence of electron flow, and some elements of circuitry are functions of submicroscopic or ionic arrangements, and so interrelated as to be essentially uncontrollable. On the other hand, tube structures separate the beam forming, power producing, and frequency control functions far better, and give far more flexibility to the circuit designer.

Multiple-Unit Devices

As a corrolary to the above difference in product control and variability, it is easy to see that tubes will continue to have an economic advantage in multifunction units, such as twin triodes or triode-pentodes, where such combinations in transistors would not prove practical. For example, if the yield of a particular tube element, when tested to the allowable customer limits, were 90%, the yield of a 2-unit tube with random variations in each element would be 0.9×0.9, or 81%. Contrast this with present transistor yields in any one spread of characteristic limits of around 25%, and a twin-unit yield would become 0.25×0.25, or about 6%.

Nuclear Radiation

A survey of the available unclassified data shows transistors to be definitely vulnerable to high-energy gamma radiation, which does not affect tubes to any measurable degree. For thermal and fast neutrons, the various reports indicate that for equal dosage, glass tubes will last about 100 times as long, and ceramic tubes about 1,000 times as long as a transistor. To quote a report of the American Society for Testing Materials:

> Thermal neutrons are not a major contributing factor to the deterioration of semiconductor devices.
>
> For the integrated doses used, gamma radiation causes surface damage which results in changes in leakage currents. Partial recovery may occur when the device is removed from the radiation field.
>
> Fast neutrons appear to cause bulk changes in materials.
>
> In reactor irradiations, silicon rectifying devices which deteriorate will probably fail in the forward direction with increases in forward resistance and break-point voltage. For the integrated doses used, germanium devices generally suffer major damage in the reverse direction only.
>
> For most applications, silicon diodes and rectifiers tested could probably be used to $10^{13}\ nv_0t$, and the silicon and germanium transistors irradiated could probably withstand $5\times10^{11}\ nv_0t$ under the conditions of test. These limits are considerably below the limits that would be caused if radiation damage on the bulk materials above is considered. The basic processes of semiconductor devices are probably disturbed by the presence of the radiation field, causing the radiation resistance of the devices tested to be considerably lower than would be expected by considering the basic properties of the semiconductor materials.[6]

An examination of another voluminous study, an Admiral Report, reveals this same ratio. "The *6L6WGB* beam power, the *5639* subminiature video amplifier, the *5840* subminiature pentode, the *5670* miniature twin tridoe, and the *5876* subminiature uhf survived the irradiation, but had control grid currents in excess of the maximum specification... The *6112* subminiature twin triode, and the *5751* miniature triode all survived the CP-5 experiment."[7] The interesting point, however, is that the transistor tolerance limit is about equal to the human tolerance limit, and that tubes are relatively unaffected at this dosage. This means that while it may be possible to evacuate personnel and protect them from an atomic explosion, it would be highly impractical to bury or shelter all electronic equipment, and if the equipment is expected to survive in operating condition, it had better use tubes.

It is interesting to note that the U.S. Army Signal Corps has become concerned about this and other tactical aspects of the use of transistors in the design of military gear, including missiles, and is publishing a comparison of its own, entitled "A Report of the Current Status and Future Trends in Electronic Tube and Transistor Electronics."

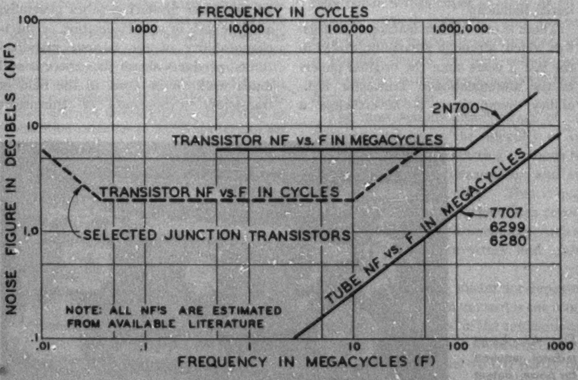


Fig. 2. Tube and transistor noise performance

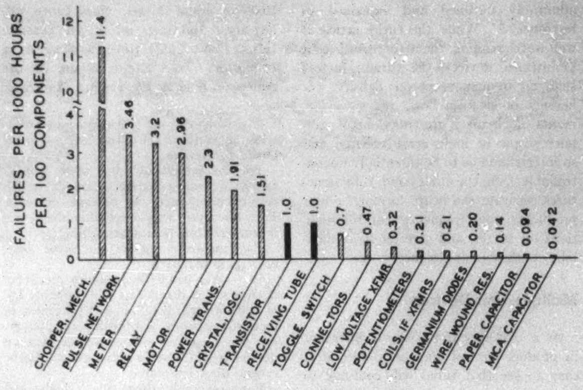


Fig. 3. Component failure rates

Additional material may be found in references 8 and 9.

Reliability

In terms of survival under equivalent operating conditions, there is very little information comparing tubes and transistors directly, except for some studies by Bell Laboratories on some of their own early types. These indicate comparable figures for their best tubes and best transistors, under practically ideal operating conditions of around 0.03% per 1,000 hours. Other less accurate figures in military gear have given about 0.1% per 1,000 hours for both.

As a practical matter, voluminous specifications, both manufacturers' and military, are available which guarantee a "floor" under allowable failure rates of tubes, and a very good probability that most lots will give better than 99% survival at 1,000 hours under maximum rated conditions, whereas such survival specifications and documented data are completely lacking for transistors. This is not to say that transistors do not have a very high potential reliability, possibly somewhat better than that of tubes, but that prediction, control, and statistical evidence of consistent survival under comparable environmental conditions are not yet available, and will probably take 10 years to accumulate. This will render impossible any calculations of system reliability or the value of such provisions as multiple channel redundancy.

A typical list of failure rates of components in a number of military equipments is shown in Fig. 3, where both tubes and transistors have been used in various portions of the equipment, as well as germanium and silicon diodes. It will be noted that in this particular set of data, the tubes and transistors came out about even at 1.0% per thousand hours. It should be noted, however, that in general it takes 3 or 4 times as many transistors to perform the same function as currently available military tube types, because of their lower gain-bandwidth factor and single-unit limitations, so that it would appear that the over-all equipment reliability using tubes is still about 3 to 4 times as good.

In fact, one example of comparable designs of a servo-amplifier required only 8 tubes in the original prototype, but wound up with 40 transistors in the redesigned version! Obviously, it will take a considerable improvement in the reliability of transistors to make up for this initial handicap.

There is one further feature of transistors which, although greatly improved in the last 2 years since the original papers in the aforementioned Transistor Reliability Symposium, is nevertheless a serious and disturbing factor, and that is simply shelf life. There is, of course, the effect of storage at high temperatures, which shows definite deterioration at 85 C for germanium,[10] and also the effect of small amounts of contaminants inside the sealed enclosure affecting the surface condition of the germanium crystal.[11] As sealing, exhaust, and gettering techniques approach that of vacuum tubes, this effect will undoubtedly be reduced, but the inherently greater sensitivity of the resistivity of the semiconducting surface to extremely minute amounts of contaminants is a problem which is not, fortunately, found in tubes.

Physical Size and Weight

While transistors obviously have the advantage over tubes on both of these counts, except for ceramic tubes, the unit itself is not the whole story. In the case of power transistors the heat sink and radiator must be many times the size of the transistor, and may well be larger than a tube of the same output rating, since the junction temperature must be kept much lower than that of the tube envelope. Similarly, in large installations involving many thousand units, the air conditioning may have to be just as large for transistors, in spite of their lower power consumption, since the allowable rise can not be as great. Heat flow and thermal calculations are fairly long and involved, so it is impossible to generalize, but this hidden factor must be borne in mind when considering tubes versus transistors from this standpoint; see Fig. 4.

Conclusions

There is ample evidence that transistors are far from displacing tubes generally, and the lack of standardization or interchangeability of the various manufacturers' products shows that there is still much work to be done in the field of transistors. Proponents of transistors

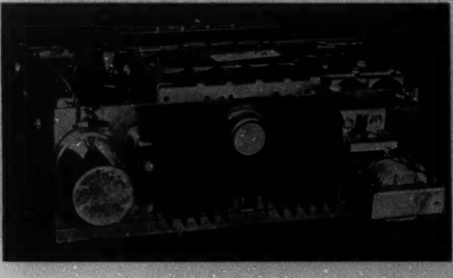


Fig. 4. Large radiator required for power output transistor

Vol
Pt.
196

make many claims for the future, but, except for some of the obvious and inherent differences in "modus operandi," tubes can claim the same opportunities for future improvements in costs via new designs and automation, and superior performance via new techniques now under development.

In the end, the equipment designer must base his decisions on the available evidence, and say, "Show me your data."

In addition to the literature cited in the paper, further material on this subject may be found in the list of references.

References

1. Factors in the Reliability of Germanium Power Transistors, **A. B. Jacobson.** *Proceedings,* Transistor Reliability Symposium, New York University Press, New York, N. Y., 1956, p. 30.

2. Transistor Reliability and Military Requirements, **S. B. Petrillo, Howard Moss.** *Ibid.,* 1958, p. 3.

3. Transients and the Transistor in Airborne Applications, **M. W. Caquelin.** *Internal publication no. CER-878,* Collins Radio Company, New York, N. Y., Dec. 1, 1958.

4. *Aviation Week,* New York, N. Y., Aug. 4, 1958.

5. On the Virtues of Separability—Tubes vs. Intrinsic Devices, **Marshall Pease.** *Microwave Journal,* Horizon House, Inc., Boston, Mass., vol. 1, no. 2, Sept.–Oct. 1958.

6. Effects of Radiation on Electronic Components: Semiconductor Devices, **M. A. Xavier, S. Nelson, A. Yefsky, A. Walters, G. J. Rotariu.** *Special Technical Publication no. 220,* "Symposium on Radiation Effects on Materials," American Society for Testing Materials, Philadelphia, Pa., vol. 2, June 20, 1957.

7. *Admiral Report no. 8,* Admiral Corporation, Chicago, Ill., phase 1, July 1957.

8. Radiation Effects in Semiconductor Devices, **Jay C. Pigg, C. C. Robinson.** *Proceedings,* Transistor Reliability Symposium, 1956, p. 77.

9. Effects of Gamma Radiation on Transistor Parameters, **Howard L. Steele, Jr.** *Ibid.,* p. 96.

10. The Reliability of Power Transistors in DC-to-DC Converters, **Bernard Reich.** *Ibid.,* p. 18.

11. *Ibid.,* p. 28.

12. The Effect of Nuclear Radiation on Selected Semiconductor Devices, **G. L. Keister, H. V. Stewart.** *Proceedings,* Institute of Radio Engineers, New York, N. Y., July 1957.

13. Evaluation of Z-2353 Tubes at 590°F Temperature, **R. N. Spear, Jr.** *Internal Publication SK*-56001-179, General Electric Company Missile and Ordnance Systems Department, Burlington, Vt., Jan. 17, 1958.

14. Electron Tubes for Critical Environment, **Walter H. Kohl, Philip Rice.** *Technical Report no. 57-434, ASTIA Document no. AD-151158,* Wright Air Development Center, Dayton, Ohio, Mar. 1958.

15. Physical Mechanisms Leading to Deterioration of Transistor Life, **G. C. Messenger.** *Transactions,* Professional Group on Electron Devices, Institute of Radio Engineers, vol. PGED-5, July 1958, p. 147.

The Empretron, a Mercury Pool Arc Tube Allowing Operation at Repetition Rates in the Kilocycle Range

K. G. HERNQVIST
NONMEMBER AIEE

F. H. NORMAN
NONMEMBER AIEE

Synopsis: The Empretron is a new mercury pool arc tube for switching applications. The arc between the mercury pool cathode and the main anode is initiated by a vacuum spark. The spark discharge occurs between two continuously replenishable electrodes, one being an electrode partly immersed in the mercury pool and the other, the starter electrode, a mercury reservoir supported above the mercury pool. Starting is accomplished by applying a high-voltage pulse to the starter electrode. The resulting spark discharge develops into an arc between the cathode and the main anode. The minimum energy necessary for arc initiation, 10^{-3} to 10^{-2} joules per start, is determined by the stray capacitance between the starter electrode and the mercury pool. Operation at repetition rates up to 5,000 cps (cycles per second) has been demonstrated. This is about one order of magnitude higher than the upper limit of 400 cps for the conventional ignitron.

Because of a materials problem resulting from the requirement that the outside of the starter electrode remain unwetted by mercury, the present tube design is limited to operation at relatively low duty cycles for the main arc. The main use of this device is therefore in pulse modulator service.

IN CERTAIN respects the mercury pool arc comes closer to the ideal of an electronically controlled switch than any other known device. Its resistance in the nonconductive state is almost infinite, withstanding voltages of hundreds of kilovolts. The anode voltage in the conductive state is very low; it can be measured in tens of volts. The mercury pool cathode is very rugged, allowing passage of thousands of amperes current. Transition from the conductive to the nonconductive state[1,2] may occur within times as short as 10^{-6} seconds. These properties and its long tube life have made the mercury pool arc a popular device for high-voltage and high-current switching applications, rectifiers, inverters, etc.

The most serious disadvantage of the mercury pool arc is the difficulty of arc initiation. Reliable starting methods require large power consumption and expensive starting circuitry. This is particularly objectionable where high repetition rates of starting are required, such as in high-frequency inverters and pulse modulators.

The most reliable starter for repeated starting that has been devised is the semiconductor starter used in the conventional ignitron tube.[3,4] Here a current is sent through a semiconductive rod partly immersed in the mercury pool. This starter requires a starting energy of about 1 joule per start and appears to be limited to repetition rates not exceeding about 400 cps.

The electrostatic-type starter[5] is considerably less power consuming. Here the starter electrode is electrically isolated from the mercury pool by a dielectric material such as glass or ceramic. A high positive voltage pulse is applied to the starter electrode for arc initiation. This device has not been of any commercial importance because of its limited life. Destruction of the dielectric material usually occurs after less than 100 hours of operation.

It is the object of this paper to describe an electrostatic starter which preserves the property of low power consumption for starting yet offers possibilities of reliability and long life. Here the arc discharge is initiated by a vacuum spark. The spark discharge occurs between two continuously replenishable electrodes, one being the mercury pool cathode and the other a mercury reservoir supported above the mercury pool. The name Empretron will be used for this device.

Description of the Principle of Operation and the Tube

Except for the case of glow-to-arc transitions it appears that the first step in the process of mercury pool arc initiation must be field electron emission from

---

Paper 60-77, recommended by the AIEE Electronics Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., January 31–February 5, 1960. Manuscript submitted October 6, 1958; made available for printing December 11, 1959.

K. G. Hernqvist and F. H. Norman are with the RCA Laboratories, Radio Corporation of America, Princeton, N. J.

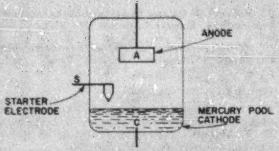


Fig. 1. Schematic diagram of tube for vacuum spark initiation of a mercury pool arc

the cathode. This is usually assumed to be true for presently known arc starting methods.[4,5] It is also corroborated by studies of the maintenance of the mercury pool arc.[6,7]

Consider a 3-element tube as shown in Fig. 1. Near the mercury pool cathode *C* is placed the starter electrode *S*. The anode *A* is placed farther away. Before starting, the space between *S*, *A*, and *C* is occupied only by neutral mercury atoms, the vapor pressure being determined by the temperature of the mercury pool and the tube walls. This pressure is about 10^{-3} mm (millimeters) of Hg at room temperature. After the arc has been established a vapor pressure of several hundred mm of Hg has been built up in the immediate vicinity of the cathode. The pressure falls off very rapidly away from the cathode, being about 10 mm of Hg at 10^{-3} cm (centimeters) from the cathode and about 10^{-3} mm of Hg at the tube walls. Furthermore a very dense plasma has been built up close to the cathode surface, a plasma capable of supplying sufficient positive space charge at the cathode surface to support electric fields of the order of 10^6 volts per cm or higher. It is the purpose of the arc starting mechanism to achieve the transition between these two states. In the Empretron, part of the energy necessary for this process is supplied by the starter circuitry and the remainder is delivered by the anode supply. The starter only supplies the seed for an arc cathode spot, the seed being a vacuum spark between *S* and *C*, only developed to the point where the anode can take over and build up the discharge to a normal arc.

The process of establishing a vacuum spark is believed to proceed as follows.[8] Consider a positive high voltage suddenly applied between *S* and *C*. The resulting high field will eventually result in field electron emission from *C*. Since the mean free path for these electrons is much longer than the separation between *S* and *C*, the electrons emitted will be accelerated and impinge on the starter electrode. Since the field emission current density is extremely high the energy input per unit area to *S* is also high, resulting in immediate vaporization of electrode material at *S*. Thus a high vapor pressure will build up around *S*, the vapor front proceeding toward *C*. Consequently new electrons emitted from *C* will encounter ionizing collisions between *C* and *S*. The ions thus formed will proceed toward the cathode and contribute to the electric field strength at the cathode. This process will now rapidly proceed to the point where ions, generated between *S* and *C* both by direct ionization and by secondary processes,[7,9] alone are capable of maintaining the high electric field at the cathode surface. Meanwhile the vapor pressure in the vicinity of the cathode surface starts building up because of the ion impact. Thus a point is eventually reached when the electric field between *C* and *A* starts to become important and a current flow to the anode builds up the main arc.

It is evident that to utilize this process for reliable arc starting the following conditions must be satisfied: 1. The spacing between *S* and *C* must be small, yet accurate and not changing in time. 2. The starter electrode must consist of an easily vaporizable material, to give a high rate of vaporization. 3. The starter electrode, as well as *C*, must consist of a continuously replenishable material. The Empretron incorporates a mercury pool arc starter built on these principles.

The tube is shown in Figs. 2 and 3. Fig. 4 shows the details of the starter assembly. The starter electrode is supported immediately above a pointed nickel electrode which is partly immersed in the mercury pool. The nickel is wet with mercury, thus keeping the point covered at all times by a film of mercury. The starter electrode itself consists of a mercury reservoir which, through the wetted capillary tubing I, communicates with the lower pointed portion of the starter electrode. This provides a mercury meniscus at the lower part of the starter electrode at all times. The mercury is prevented from dripping out of the reservoir by the surface tension at the meniscus, provided that *T* has a sufficiently small diameter. Also the outside of the starter electrode is prevented from being wet by mercury by the wetting shield or by special surface treatment of

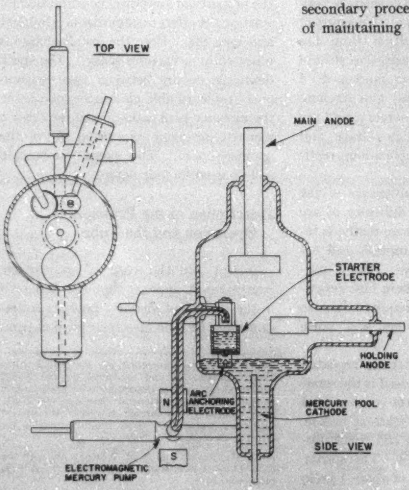


Fig. 2 (left). Empretron, using electromagnetic pump

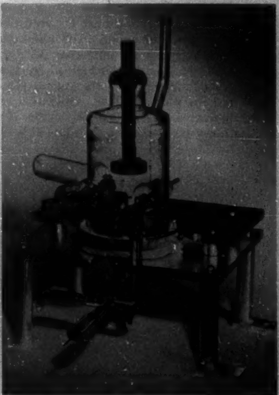


Fig. 3 (right). Empretron

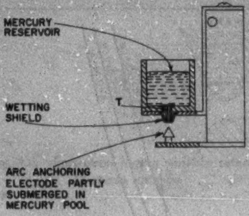


Fig. 4. Starter assembly

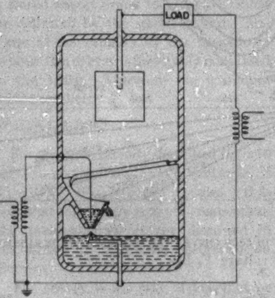


Fig. 5. Empreton using condensed mercury for reservoir

the starter electrode. The mercury in the reservoir is constantly replenished either by allowing mercury condensed on the walls to flow into the reservoir as shown in Fig. 5 or by means of an electromagnetic pump as shown in Fig. 2. In the first method, Fig. 5, the condensed mercury is pulled down by gravity and collected by the gutter slightly tilted toward the starter. In the second method, Fig. 2, mercury is pumped up from the mercury pool.

Operation of the tube consists of applying a high positive voltage to S. After a time which is determined by the field strength between S and C a vacuum spark breakdown occurs between C and the mercury meniscus at the lower part of S. The arc is established between C and the main anode or the holding anode.

Maximum Prebreakdown Energy

The prebreakdown energy is the energy required to raise the vapor pressure between the starter electrode and the cathode to a value high enough for breakdown to occur. It will be shown in this section that this energy is much less than the energy which is inevitably used in the breakdown process itself.

As discussed in the preceding section the process of establishing a vacuum spark involves a front of vapor proceeding from the starter as well as field electron emission from the cathode. The value of vapor density necessary for breakdown is very difficult to estimate. However it is known that when a uniform vapor pressure in the region between C and S has attained a certain value, breakdown in accord with the Paschen curve will occur. The energy necessary for vaporization to obtain this pressure represents the upper limit of required prebreakdown energy. Fig. 6 shows a typical Paschen curve. For the case of interest, the pressure p_0 before application of starter voltage, as well as the spacing d between C and S, is very small. Thus the line $p_0d=\text{constant}$ lies to the left of the minimum of the Paschen curve. Assume that a voltage V_s is suddenly applied between S and C. This may represent the point $(V_s;p_0d)$ in Fig. 6 which lies outside of the breakdown region. As evaporation of the starter electrode sets in, the pressure rises in the region between S and C. When the point $(V_s;p_1d)$ in Fig. 6 has been reached, breakdown is certain to occur. The maximum prebreakdown energy is the energy required to raise the vapor pressure between S and C from p_0 to p_1. Because of the high density of energy input to S other sources of heat loss such as heat conduction may be neglected.

The Paschen curve for mercury is shown in Fig. 7. The part to the right of the minimum is taken from data by Llewellyn Jones and Galloway,[10] the part to the left from Guseva and Klarfeld.[11] For high voltages and low values of pd the curve can be expressed as[11]

$$V=\frac{210}{(pd)^2} \tag{1}$$

where V is in volts, p in mm of Hg, and d in cm.

The energy required to raise the temperature of mercury from room temperature to the boiling point and for evaporation[12] is $\eta=1.13\times10^{-19}$ joules/atom. Assuming a plane parallel geometry of area A cm^2 between S and C the maximum prebreakdown energy in joules can be written as

$$W_{PB_{\max}}=\eta\times3.56\times10^{16}(p_1-p_0)Ad \tag{2}$$

Since $p_0<<p_1$ and inserting equation 1

$$W_{PB_{\max}}=5.83\times10^{-2}\frac{A}{V_s} \tag{3}$$

As an example take $A=0.1$ cm^2 and $V_s=10^4$ volts yielding

$$W_{PB_{\max}}=5.8\times10^{-5}\text{ joules}$$

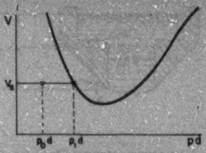


Fig. 6. Typical Paschen curve

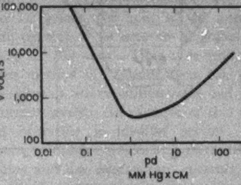


Fig 7. Paschen curve for mercury vapor

Breakdown measurements using series resistance indicate that the actual prebreakdown energy is much lower than $W_{PB_{\max}}$. This is also corroborated by the measurements of Boyle, Kisliuk, and Germer.[8]

Breakdown Time Lag

It has been found[5] that breakdown between a mercury cathode and an electrode occurs at much lower average electric fields (V_s/d) than is necessary for appreciable field emission currents. The reason for this is the mechanical disturbance in the mercury surface caused by the electric field. Under the influence of the electric field, microscopic deformation of the liquid surface into sharp points results in much higher local fields at the cathode surface than at a rigid surface. These sharp points are the centers for field electron emission. There is a certain time lag τ between the application of the field and the formation of these sharp points. This time τ proceeds and is usually much longer than the process of raising the vapor pressure described in the previous section.

Both theoretical and experimental studies of the time lag τ have been done. Warmholz[5] has measured the time lag in breakdown between a mercury cathode and a mercury electrode. These results have been verified by studies on the Empretron starter. It appears that for field strengths $E_s=V_s/d$ exceeding 250 kv/cm the time lag τ in μsec (microseconds) is approximately given by the empirical expression

$$\tau=280\times Exp\left(-\frac{E_s}{422}\right) \tag{4}$$

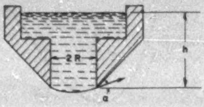


Fig. 8 (left). Schematic drawing of starter electrode

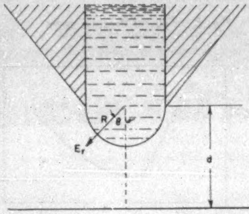


Fig. 9 (left). Schematic drawing showing shape of mercury meniscus at stability limit

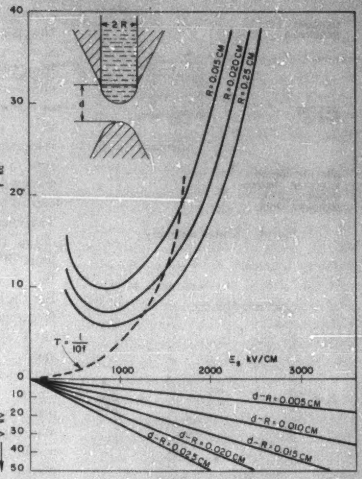


Fig. 10 (right). Frequency limits for Empretron

where E_s is measured in kv/cm. Thus a field of 422 kv/cm gives a time lag of about 100 μsec.

Meniscus Stability

Of primary importance for the operation of the Empretron is the stability of the mercury meniscus at the lower portion of the starter electrode. Fig. 8 shows schematically a cross section of the starter electrode. It will be assumed that the inside of the narrow cylindrical feed tube, with a diameter of $2R$, is well wetted by the mercury but that the outside of the starter electrode is unwetted. The head of mercury h is then supported by the surface tension at the meniscus at the lower part of the feed tube. The equilibrium condition when no potential is applied between S and C is

$$9.81\rho\ \pi R^2\ h=2\pi\ R\sigma\sin\ \alpha \qquad (5)$$

or

$$h=\frac{2\sigma\ \sin\ \alpha}{9.81\ \rho\ R} \qquad (6)$$

where h and R are expressed in meters, $\sigma=$ surface tension in newtons/m (per meter) and $\rho=$ specific weight in kg/m^3. The angle α is defined in Fig. 8. The stability criterion is $\alpha<90$ degrees. For mercury at room temperature $\sigma=0.47$ newtons/m and $\rho=13.5\times10^3$ kg/m^3. As an example take $R=10^{-4}$ m and $h=10^{-2}$ m. In this case, according to equation 6, $\alpha=8$ degrees.

In the case when a high voltage is applied between S and C the additional electrostatic force F on the mercury meniscus must be considered. This force adds to the gravitational force of equation 5. In analyzing this case the following assumptions will be made: 1. $F>>9.81\ \rho\pi R^2h$. This can always be satisfied by making h small enough. 2. The meniscus shape is spherical. The experimental observations indicate that this is a good approximation. 3. The normal component of the electric field at the meniscus surface varies as $E_r=E_s\cos\theta$ where E_s is the maximum field and θ is defined in Fig. 9.

For the case of most interest the voltage is applied to the starter periodically at a repetition rate f starts per second. The voltage between S and C is a train of pulses of pulse duration τ, as seen in equation 4, and frequency f. It will be assumed that the inertia of the mercury column is large enough so that the average electrostatic force is given by

$$F=F_0\ \tau\ f \qquad (7)$$

where F_0 is the force for a d-c potential applied between S and C. With the above assumptions F_0 for the stability limit $\alpha=\pi/2$ is given by integration of the force over the meniscus surface.

$$F_0=\int_0^{\pi/2}\pi\ \epsilon_0\ E_s^2R^2\cos^2\theta\ \sin\theta\ d\theta=\frac{\pi}{4}\epsilon_0E_s^2R^2 \qquad (8)$$

For the limiting case, $F=2\pi\ R\sigma$ or from equations 4, 7, and 8

$$Rf\ E_s^2\times Exp\left(-\frac{E_s}{422}\right)=1.52\times10^7 \qquad (9)$$

where R is in cm, f in sec^{-1}, and E_s in kv/cm. Equation 9 represents the criterion for meniscus stability. It is plotted in Fig. 10 with R as a parameter. The region of meniscus stability is to the right of each curve. E_s is related to the applied voltage V_s by

$$E_s=\frac{V_s}{d-R} \qquad (10)$$

This relation is plotted in the lower part of Fig. 10. As an example of the use of Fig. 10, assume a case where $f=20$ kilocycles, $R=6$ mils, $d=10$ mils. According to Fig. 10 this will require a starting voltage V_s larger than 18 kv for stable operation.

A further restriction on the operating range is set by the practical requirement that the time lag τ must be short compared to the period f^{-1}. If one makes the arbitrary assumption at $\tau\leq0.1\ f^{-1}$ equation 4 yields

$$f\leq357\times Exp\left(\frac{E_s}{422}\right) \qquad (11)$$

This requirement limits the operating parameters to the region to the right of the dotted curve in Fig. 10.

Arc Initiation Energy

As discussed in a foregoing section, the energy necessary for establishing an arc discharge consists of two parts. One part, which will be called the arc initiation energy W_i, is delivered by the starter. The other part is delivered by the anode circuit and is of no consequence for the starter circuit.

Quite independently of the physical process of arc initiation, there exists a lower limit on W_i which is determined by the stray capacitance C_s between the

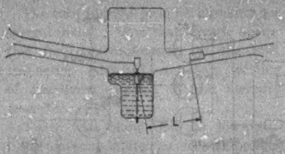


Fig. 11. Empretron having variable cathode to anode distance

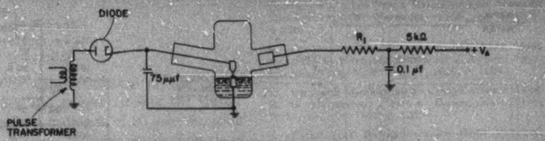


Fig. 12. Diagram showing test circuit for Empretron

starter electrode *S* and cathode *C*. When a voltage V_s is applied between *S* and *C*, a stored energy equal to 1/2 $C_sV_s^2$ must be delivered by the starter circuit. Because of the rapidity of the breakdown process this energy is inevitably consumed in the spark. Therefore the minimum initiation energy is

$$W_{t_{\min}} = 1/2\ C_sV_s^2$$

For the tube already described, $C_s \approx 10$ μμf (micromicrofarad). For example, if $V_s = 10$ kv, $W_{t_{\min}} = 5 \times 10^{-4}$ joules. It is seen that this energy is much greater than the prebreakdown energy discussed in the section on Maximum Prebreakdown Energy.

The energy $W_{t_{\min}}$ may or may not be sufficient for reliable starting depending on the operating conditions. An important factor is the electric field which the potential between the anode and the cathode exerts at the point of the spark breakdown. The necessary degree of development of the seed for the arc cathode is difficult to estimate theoretically and has to be determined experimentally for each particular tube design. As an example of the effect of the anode potential and anode-to-cathode distance on the misfiring rate, tests were performed on the tube structure shown in Fig. 11. The distance *L* between the main anode and the mercury pool could be varied. Measurements were performed at $L = 2$, 4.5, and 9 centimeters. The test circuit is shown in Fig. 12. High-voltage pulses were applied to the starter electrode at a 60-cps repetition rate. The pulse voltage V_s was 18 kv which was sufficient to assure a 60-cps breakdown rate for a spacing $d = 0.010$ inch. The capacitance C_s was 75 μμf, resulting in an energy 1/2 $C_sV_s^2 = 1.22 \times 10^{-2}$ joules per start. The anode supply voltage V_A and the anode current-limiting resistor R_1 were variable and adjusted to yield a peak anode current of 10 amp (amperes). The number of anode firings per second was measured by means of a counter coupled to the anode. For each spacing *L*, the firing rate was measured as a function of anode voltage V_A. The results are shown in Fig. 13. As seen, the misfiring rate increases with *L* and decreases with V_A. The geometry is unfortunately too complicated to give any significance to the dependance of firing rate on V_A/L.

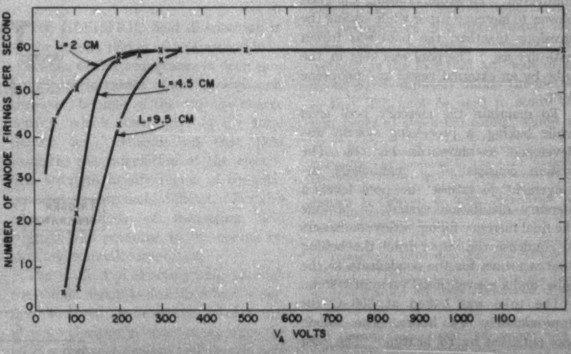


Fig. 13. Diagram showing anode firing rate as a function of anode voltage with cathode to anode distance L as a parameter. 60 corresponds to 100% firing

Mercury Consumption in the Starter Electrode

A factor of great practical importance in the operation of the Empretron is the amount of mercury evaporated from the starter electrode. The demand on mercury replenishment to the starter reservoir determines the type of replenishment method to be used.

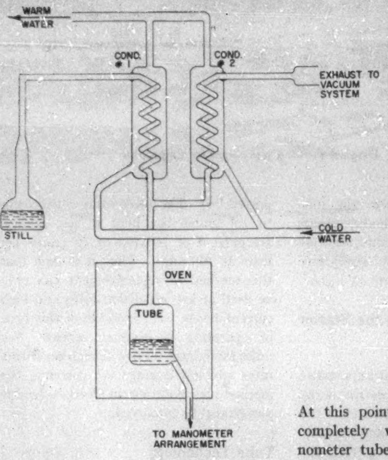
From the evaporation energy, discussed in a preceding section, the evaporation rate is found to be 10 grams per hour for a power input of 1 watt to the starter. This may be a representative figure for the evaporation rate of a tube operating with repetition rates in the kilocycle region.

For the replenishment method shown in Fig. 5, mercury evaporated by the action of the main arc is condensed on the tube walls and allowed to flow into the starter reservoir. The evaporation rate from the mercury pool cathode varies greatly depending on the operational conditions. Evaporation rates in the range 0.2 to 36 milligrams per amp of arc current have been reported.[13,14] The lower value corresponds to 7.2 grams per hour at an average arc current of 10 amp or 72 grams per hour at 100 amp. Thus it is seen that this method of replenishment can only be used at low repetition rates and high current levels. An example of this type of operating is inverter service. For pulse modulator service at high repetition rates and low average arc currents, the method incorporating an electromagnetic pump must be employed.

Tube Technology

The requirement that the upper starter electrode should not be wet on the lower outside surfaces has presented a materials problem. A variety of materials and shielding coatings, including sapphire, was tried before arriving at the present structure. The severe environmental conditions such as local heating by electron impact, ion bombardment, etc., peel off or reduce most materials to expose clean metal which promptly wets with mercury. A metal cup is used for its good heat conductivity.

The present starter reservoir is made of molybdenum which is then carburized. This surface layer of molybdenum carbide has extreme resistivity to mercury wetting. The carburizing process consists of raising the cup to 1,500 C (degrees


WARM WATER
COND. # 1
COND. # 2
EXHAUST TO VACUUM SYSTEM
STILL
COLD WATER
OVEN
TUBE
TO MANOMETER ARRANGEMENT

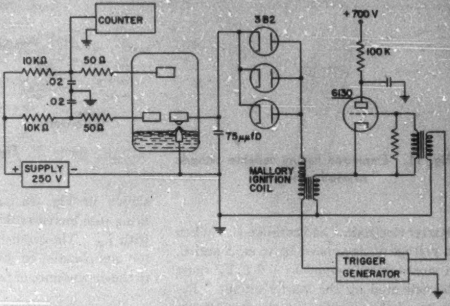

COUNTER
10KΩ
50Ω
.02
.02
10KΩ
50Ω
+ SUPPLY 250 V −
75μμfD
3B2
+700 V
100 K
6130
MALLORY IGNITION COIL
TRIGGER GENERATOR

Fig. 14 (left). Schematic diagram showing tube processing apparatus

Fig. 15 (above). Life test circuit

centigrade) for several hours by r-f (radio-frequency) heating in an atmosphere of toluene bubbled hydrogen. The capillary tube, a nickel tube of 0.006-inch inside diameter is inserted in the molybdenum cup after carburizing.

The inside of the nickel capillary and the pointed cathode electrode, after a chemical cleaning, are prewet with mercury before assembly of the tube.

Mercury-wetted surfaces are maintained with great difficulty, exposed in air or to high-temperature baking cycles during tube processing. They stay wet indefinitely, however, when stored under a pool of mercury. For this reason the electrodes, after mounting in place, were submerged in a mercury pool which was maintained until processing was completed. To avoid exposing the glass blower to mercury vapor while sealing the envelope together, the pool was frozen with dry ice. The pool was kept in the tube by an external vessel and neoprene hose.

To maintain the mercury pool level while baking, a processing system was developed, as shown in Fig. 14. The system consists of a manometer arrangement to adjust the pool level, a mercury distillation system to provide the final mercury filling, series condensers to condense the vapor from the boiling pool, a return for the condensate to the tube, and a conventional vacuum system.

The tube was baked at 360 C for approximately 24 hours. The anode was outgased by r-f heating. The pool was lowered and the starter was operated for 1/2 hour, anode potential applied. At this point the protective pool was completely withdrawn and the manometer tube sealed off. The still was operated until the proper level was reached. Seal-off completed the tube processing.

Life testing was done in the circuit shown in Fig. 15. Test parameters were as follows:

Starter

Firing Rate . . . . $f=60$ pps
Pulse Length . . . . $\tau=100$ μsec
Pulse Voltage . . . . $V_s=15$ kv
Capacitance . . . . $C_s=75$ $\mu\mu$f

Anodes: Main and Holding
Voltage . . . . $V_A=250$ volts
Peak current = 4 amp
Duty cycle = 0.001

The test was interrupted at 650 hours. No outside wetting of the starter reservoir was apparent, and the anode firing followed the starter perfectly.

An additional test was performed with the parameters shown in the following tables:

Starter

Firing rate = 60 pps
Pulse length = 100 μsec
Pulse voltage = 15 kv

	Anodes	
	Holding	**Main**
Voltage	250 volts	249 volts
Peak current	4 amp	12 amp
Duty cycle	0.001	0.400

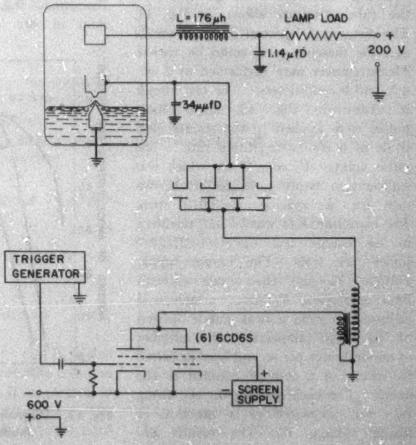

L = 176μh
LAMP LOAD
+ 200 V
1.14μfD
34μμfD
TRIGGER GENERATOR
(6) 6CD6S
+
−
SCREEN SUPPLY
− 600 V +

Fig. 16. High-frequency circuit

The tube showed good performance for several hours. After that, erratic operation forced abandonment of the test. Inspection revealed complete wetting of the lower portion of the mercury reservoir.

These tests show that for low-duty-cycle pulse-modulator service, the present structure will provide reasonably long tube life. For long duty cycles however a materials problem still exists.

High-Frequency Tests

The Empretron has been tested by H. Cooke and O. Dow of the RCA Laboratories at firing rates of up to 5,000 pps. The circuit used is shown in Fig. 16. The tube had the following parameters:

Starter
Pulse rate = 5 kc burst for 20 millisec per second
Capacitance = 37 $\mu\mu$f
Voltage = 10 kv
Pulse width = 10 μsec
Spacing $d = 0.005$ inch

Anode
Voltage = 200 volts
Peak current = 4 amp
Duty cycle = 0.03

To avoid overloading of the power supplies and trigger circuits the starter was pulsed in bursts of 20 μsec duration each second. Reliable anode firing was demonstrated up to 5,000 pps. Limitations of anode heat dissipation in the experimental tube prevented operation at higher repetition rates.

References

1. Über die Löschung des Lichtbogenbrennflecks an einer Quecksilberkathode, **G. Mierdel.** *Zeitschrift fuer Technische Physik*, Leipzig, Germany, vol. 17, 1936, p. 452.

2. Über rasche Vorgänge im Quecksilbervakuumbogen, **O. Engelbrecht.** *Archiv fur Elektrotechnik*, Berlin, Germany, vol. 36, 1942, p. 515.

3. A New Method for Initiating the Cathode of an Arc, **J. Slepian, L. R. Ludwig.** *AIEE Transactions*, vol. 52, June 1933, pp. 693–700.

4. Time Lag of an Ignitron, **N. Warmoltz.** *Phillips Research Reports*, Eindhoven, Netherlands, vol. 6, 1951, p. 383.

5. The Time-Lag in Starting a Low-Pressure Arc on a Mercury or Gallium Cathode in Connection with Field Emission and Surface Deformation, **N. Warmoltz.** *Ibid.*, vol. 2, 1947, p. 426.

6. Discharge Mechanism of Mercury Pool Arcs, **K. G. Hernqvist.** *Journal of Applied Physics*, New York, N. Y., vol. 27, 1956, p. 1226.

7. Emission Mechanism of Cold-Cathode Arcs, **K. G. Hernqvist.** *Physical Review*, New York, N. Y., vol. 109, 1958, p. 636.

8. Electrical Breakdown in High Vacuum, **W. S. Boyle, P. Kisliuk, L. H. Germer.** *Journal of Applied Physics*, vol. 26, 1955, p. 730.

9. The Excitation Theory of Arcs with Evaporating Cathodes, **A. von Engel, A. E. Robson.** *Proceedings*, Royal Society of London, London, England, series A, vol. 242, 1957, p. 217.

10. The Sparking Potential of Mercury Vapor, **F. Llewellyn Jones, W. R. Galloway.** *Ibid.*, vol. 50, 1938, p. 207.

11. Ignition Voltage of the Discharge in Mercury Vapours, **L. G. Guseva, B. N. Klarfeld.** *Zhurnal Tekhnicheskoi Fiziki*, Moscow, USSR, vol. 24, 1954, p. 1168.

12. Heat Capacity of Liquid Mercury Between 0° and 450°C; Calculation of Certain Thermodynamic Properties of the Saturated Liquid and Vapor, **T. B. Douglas, A. F. Ball, D. C. Ginnings.** *Journal of Research*, National Bureau of Standards, Washington, D. C., vol. 46, 1951, p. 334.

13. The Rate of Vaporization of Mercury from an Anchored Cathode Spot, **L. Tonks.** *Physical Review*, vol. 54, 1938, p. 634.

14. Die Verdampfung an der Kathode des Hg-Bogens, **J. V. Issendorf.** *Physikalische Zeitschrift*, Leipzig, Germany, vol. 29, 1928, p. 857.

Potentials in D-C Corona Fields

GAYLORD W. PENNEY
FELLOW AIEE

RICHARD E. MATICK
NONMEMBER AIEE

Synopsis: A probe method has been devised for measuring potentials in d-c corona fields which are uniform for a short distance in one direction. In unipolar corona, a probe supported by perfect insulators assumes a potential sufficiently above the space so that no more ions can reach the probe. For a cylindrical probe with its axis along an equipotential, this excess or "error potential" is proportional to the diameter of the cylinder. To minimize this error a fine wire is used as the probe. A frame is designed which both supports the wire and shields it from fields other than the field being measured. The fine wire is electrically connected to an electrostatic voltmeter, and the frame to a source of variable potential. A potential is measured by slowly increasing the potential of the frame and recording both frame and wire potential. Under proper conditions, the space potential is the point where wire and frame potentials are equal.

Potentials have been measured as a function of position in a corona field produced by a row of wires between parallel grounded planes. Potentials can be measured quite precisely so that gradients can be determined with reasonable accuracy; they can be measured close to the passive electrode which is the region of greatest importance in electrostatic precipitation. The method can also be used in abnormal corona where there are both positive and negative ions.

A method is outlined for calculating the potential distribution when no space charge (corona current) is present in the geometry consisting of wires between parallel planes, and results are presented. Curves are presented which show how the potential and electric field distribution changes as the corona current is changed under various conditions.

THE ELECTRIC field distribution in the corona region of a precipitator is, of course, of first importance in determining the performance of the precipitator. However, because of the way the charge density, which contributes to the total electric field, depends on that field strength, the calculation of the electric field strength in any region of irregular geometry is extremely difficult. For this reason, methods of measuring field strength and potential in the corona region are especially important.

The method of shooting balls through the corona region as described by Cooperman[1] required the measurement of charge at two velocities in order to separate the effects of charge density and electric field strength. This is difficult experimentally and it led to a considerable uncertainty in results. The method of measuring charge acquired by a ball falling through the corona as used by Penney and Hewitt[2] gave a better measurement of field strength but considerable uncertainty as to the exact position of the field strength measured. This was especially true in regions where the field strength increases in the direction of the field so that the ball is deflected into an increasingly stronger field, which is normally the condition near the dust-collecting electrode of usual precipitators. For this reason the ball-dropping technique is not suitable in the region which is of greatest importance to the performance of a precipitator. Another serious weakness of charged ball methods is the fact that any ions of the wrong sign completely nullify the method so that it cannot be used to study certain types of abnormal corona which are

---

Paper 60-33, recommended by the AIEE Electronics Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., January 31–February 5, 1960. Manuscript submitted October 16, 1958; made available for printing November 19, 1959.

Gaylord W. Penney is with the Carnegie Institute of Technology, Pittsburgh, Pa., and Richard E. Matick, formerly with the Carnegie Institute of Technology, Pittsburgh, Pa., is with the International Business Machines Corporation, Poughkeepsie, N. Y.

This investigation was supported by the Aerotec and Thermix Corporations. The authors wish to acknowledge the suggestions and encouragement of E. R. H. Pegg of the Aerotec Corporation and Sidney Orem, Jr., of the Thermix Corporation.

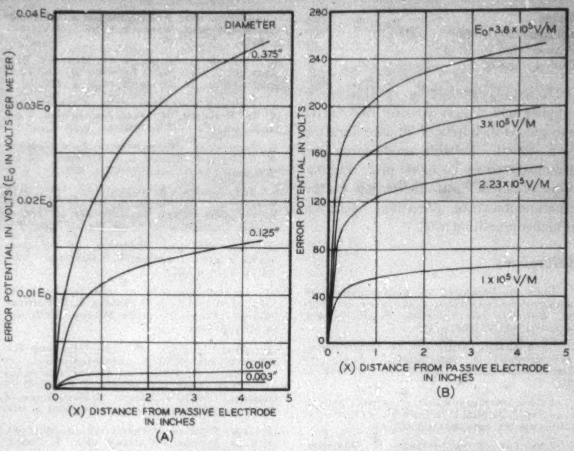


Fig. 1 (above). Calculated error potentials. A—For 4-wire diameters. B—For a 0.003-inch-diameter wire

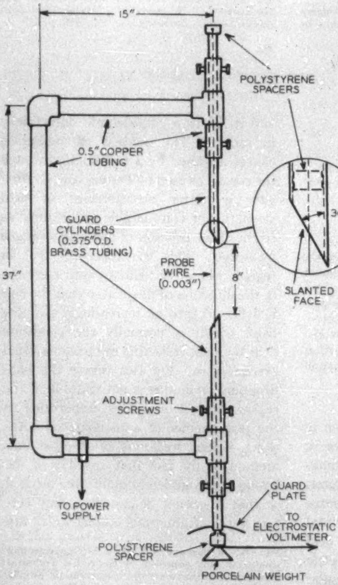


Fig. 2 (left). Probe and supporting frame

very detrimental to the performance of a precipitator.

A probe method has the advantage that there is no uncertainty in the position of the measurement. Probe methods were developed by Langmuir and others for for measuring potentials in low-pressure discharges where the mean free path for electrons or ions is large so that uncertainties were primarily due to the energies of the individual ions. In corona at atmospheric pressure the kinetic energy of the individual ions is not large. In this case the major difficulty is due to the fact that there are ions of only one sign which are driven along lines of force so that they charge any insulated probe to a potential higher than that of the space if the probe were not there. This paper describes a method of minimizing this error so that the potential of a point in space can be determined with considerable accuracy. Potential is a scalar quantity so that, once the space potential has been determined, it is a simple matter to obtain the electric field as the gradient of potential. This probe technique can be used only for measurements of d-c corona; it cannot be used in a-c corona.

General Discussion

The probe to be described requires that the potential to be measured must be uniform for an appreciable distance in one dimension and that the probe be oriented so that its axis extends in the direction along which the potential is uniform. If the probe is a cylinder of uniform diameter, the charge per unit length which it acquires due to the movement of ions of the same sign will be proportional to the diameter and proportional to the field strength. If such a long cylinder is parallel to a grounded plane, the potential which such a probe attains, in excess of the potential of the space if the probe were absent, is approximately proportional to the charge per unit length. This excess or error potential is thus proportional to the diameter of the probe and to the maximum field strength in the region to be measured. Since the field being measured must not be changed, the practical method of minimizing the error potential is to use a small probe. This requirement indicates that the probe should be a fine wire. But if such a fine wire is brought out from a high-potential region and connected to an electrostatic meter, its potential may be influenced by the distortion of equipotential surfaces near the ends of the apparatus; this is because the probe will charge to the largest magnitude of potential through which its center line passes(plus the small error potential, of course). Also, it will lose charge by corona along any length located in a low-potential region. Since the wire can receive ions from the field at a very limited rate, any loss of charge or leakage current gives a reduction in potential and so must be eliminated. The probe then consists of a fine wire supported by a frame which forms a shield to protect the wire from extraneous potential regions through which it must pass. The frame also provides for a transition to a large lead going to an electrostatic voltmeter and with the potential of the frame controlled so that the presence of the frame controlled so that the presence of the frame does not disturb the potential of the fine wire.

Error Potential

The error potential is the difference between the magnitude of the potential of the probe and that of the corresponding line in space if the probe were not there. This potential difference or error is due to the charge which the probe receives from the unipolar ionized field. An exact calculation of this error potential would be very difficult because of the nonlinear nature of the problem. However, with the following simplifying assumptions a solution is readily obtained. It will be assumed that:

1. The probe is a wire of infinite length and is parallel to an infinite grounded plane.
2. The field intensity is the same as that existing at the center location of the cylinder before the cylinder is introduced, and is produced by an infinite plane electrode located at a large but finite distance from the wire; also there are very few ions present so their space charge has a negligible effect on the field.
3. The charge on the wire is equilibrium charge obtained in an initially uniform field containing a small number of ions and with the wire located a large distance from any electrode.

These assumptions become better and better approximations to the actual case as the wire diameter is decreased. With these assumptions, the potential of the wire can be readily calculated by the superposition of the field due to the charge of the wire, on the uniform field between the planes. The equilibrium charge acquired by the wire is independent of the charge density and it is easily shown by a method similar to White's[3] that the charge per unit length ρ_l is

$$\rho_l = 4\pi a \epsilon_0 E_0 \qquad (1)$$

rationalized meter-kilogram-second units, where

ϵ_0 = permittivity of space, 8.85×10^{-12} farads per meter
a = wire radius, meters
E_0 = field strength, volts per meter

The potential of this wire above the corresponding line in space if the wire were not there (error potential) is plotted in Fig. 1(A) as a function of distance from the grounded electrode for four different wire sizes. These curves show the importance of using a fine wire as a probe. In applying these curves for error potential to actual corona, there are some situations where the approximation will not be good. If the probe is within a few diameters of any other electrode the field around the probe will be distorted so that equation 1 does not represent the charge on the probe. Another difficulty is due to the fact that the probe disturbs the distribution of ions between the probe and the passive electrode. However, in order to minimize the error potential a fine wire probe is required and as the probe diameter is decreased the calculated error potential becomes a better approximation to the actual situation.

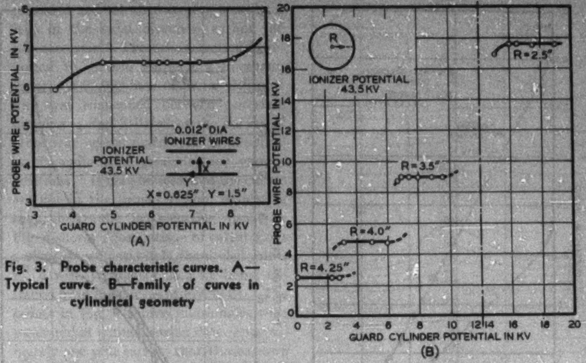


Fig. 3. Probe characteristic curves. A—Typical curve. B—Family of curves in cylindrical geometry

A wire of 0.003-inch diameter has been used as a probe. In Fig. 1(B) the error potential in volts for a wire 0.003 inch in diameter is plotted as a function of distance from the grounded plane for four values of field intensity. Since the foregoing assumptions used for calculations become better approximations as the wire diameter is decreased, the curves of Fig. 1(B) should give an estimate of the actual error potential. The error potential for a 0.003-inch-diameter wire is usually negligible. For example, the curve for 3×10^6 volts per meter shows an error potential of 165 volts at a distance of one inch from the grounded electrode. The potential at this point would be 7,620 volts, so the error potential is then 2.16%. Thus, for precise work, a correction for error potential can be approximated; see reference 4 for a more complete discussion of this error potential.

Technique for Obtaining Correct Probe Potential

Considerations of error potential dictate the use of a fine wire as a probe. If this fine wire is supported by the guard cylinders and frame as shown in Fig. 2, the potential of the guard cylinder must be controlled so that the fine wire can assume its own undisturbed potential. If the magnitude of the potential of the guard cylinder is sufficiently lower than that of the wire, there will be a corona from wire to guard cylinder, which will lower the magnitude of the wire potential. As the magnitude of the potential of the guard cylinder is increased, a point is reached where this corona from wire to guard cylinder ceases, so there is no leakage from the wire which then can attain its own potential. Under proper conditions, as the magnitude of the potential of the guard is further increased, the wire potential will remain constant. In this case the probe wire potential will remain constant until the magnitude of the guard potential is sufficiently larger than the wire to direct ions to the wire and hence increase the magnitude of its potential. Thus, the space potential is the potential of the wire in the region where wire potential is independent of guard cylinder potential except, of course, for the small error potential. An electrostatic voltmeter is used to measure the wire potential. Since it has a high insulation resistance it draws essentially no current.

The design of the guard cylinders as shown in Fig. 2 was arrived at experimentally. As mentioned above, in the flat region of the curve of wire versus guard cylinder potential the guard cylinder has no influence on the wire potential. It was found experimentally, on special apparatus, that a square-ended cylinder (i.e., one which is cut off perpendicular to its axis) influenced the wire potential when the probe was near the grounded plane, so the curve is not really flat in the desired region.

Although the increase in magnitude of the wire potential effected by the guard cylinder has been measured and found to be small (on the order of the error potential) it leads to difficulties in defining the true space potential as well as intro-

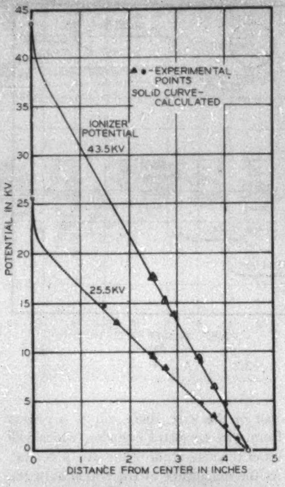


Fig. 4. Comparison of calculated and measured potentials for a 0.0063-inch-diameter wire in a 9-inch-diameter cylinder

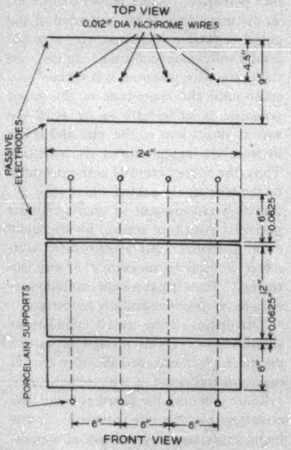


Fig. 5. Arrangement of wires between two parallel-grounded plates

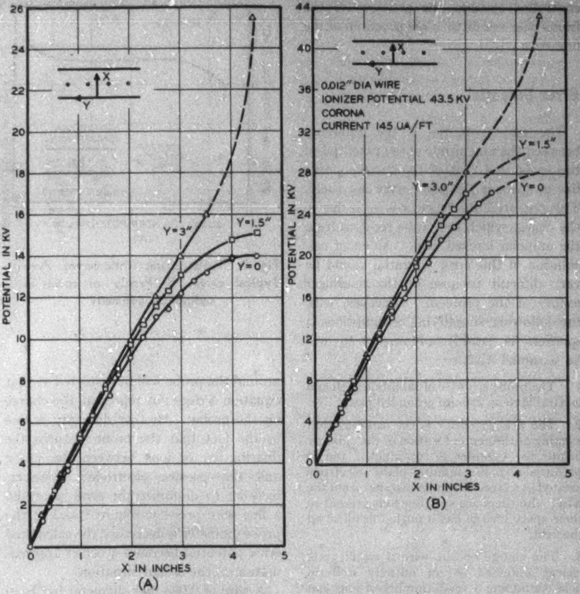


Fig. 6. Measured potential distribution in geometry of Fig. 5. Ionizer potentials in kv: A—25.5 B—43.5

ducing an additive error. It was further found experimentally that when the probe was cut on an angle as shown in Fig. 2 and with the slanted face facing the passive electrode, no influence of cylinder on wire potential could be measured in the prescribed region, while if the slanted face was turned 180 degrees, i.e., facing the active electrode (ionizer wires) the influence of guard cylinder on wire potential was much greater than that of the square-end cylinder. Experimental investigations indicated that this influence or "end effect" of the square-ended guard cylinder on the wire potential, in the region where the probe potential should be independent of guard cylinder potential, is due to a three-dimensional distortion of equipotential surfaces which causes the wire to charge slowly to a magnitude of potential above the desired value. (A more complete discussion of the end effect of the guard cylinders is given in reference 5.) It suffices to say that satisfactory results were obtained with the probe as shown in Fig. 2, and this was the design used.

The guard plate shown in Fig. 2 attached to the guard cylinder, near the point where the electrostatic voltmeter lead attaches to the probe wire, was necessary to prevent stray ions from accumulating on the connection, which would have increased the voltmeter reading. The polystyrene spacer shown at the top of the probe cylinder was made so that the probe wire was completely enclosed, thus preventing any stray ions near the top from reaching the wire.

In all apparatus used, grounded shields were attached near the bottom of the apparatus for the reason just described, as well as for keeping stray ions away from the electrostatic voltmeter itself.

In view of the foregoing, the space potential can be defined as the potential at which the wire and guard cylinder potentials are equal. A typical curve showing the variation of probe wire potential with the potential of the probe frame is shown in Fig. 3(A). The potentials shown on the curve as well as all others in this paper are actually negative but are not designated as such.

A probe wire below its correct potential can always receive ions from the field. However, a probe above its true potential cannot lose any charge until the difference in potential is so great that a loss of charge occurs by corona or other form of leakage from the wire. A curve such as that in Fig. 3(A) can be obtained in either of two ways. The magnitude of guard cylinder potential can be started at a low value and gradually increased, or it can be started at a high value and gradually decreased; in either case, precautions must be taken because of the capacitance between wire and guard cylinder. If the magnitude of the guard

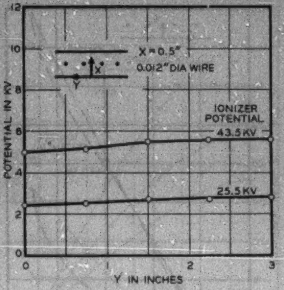


Fig. 7. Potentials measured along a line 1/2 inch from passive electrode

cylinder potential is gradually increased, the wire should be slightly discharged by means of a large resistance to ground to drain off the excess charge resulting from the capacitance effect. This should be done after each change of guard cylinder potential. Otherwise, the flat region will not appear to be flat but will have a definite slope, yielding erroneous results.

If the curve is obtained by decreasing the magnitude of the cylinder potential, the same procedure should be used until a point is reached just where the flat region begins. Then, as the magnitude of cylinder potential is further decreased, the capacitance effect between wire and cylinder will cause the magnitude of the wire potential to decrease momentarily, but it will immediately charge back to the proper value. Thus, the flat region will be obtained without the need of externally discharging the wire. Both of the methods were used with satisfactory results throughout the experimental work.

For similar reasons, both potential sources must be reasonably well regulated since any momentary increase in potential imparts an increased charge to the wire which cannot escape, thus giving the wire a potential whose magnitude is above the value corresponding to the steady potential of the field.

Tests to Prove Accuracy of Probe Method

To verify the use of the probe for measuring potentials, measurements were made using a concentric wire and cylinder. The potential distribution for this geometry can be accurately calculated for comparison with measured values.[5] The cylindrical apparatus consisted of a 9-inch-diameter cylinder, 1 foot long, and two 9-inch-diameter end sections each 6 inches long. The two end sections are grounded and are used to give a uniform field in the axial direction within the 1-foot section. The 1-foot section is returned to ground through a microammeter. A 0.0063-inch-diameter nichrome wire was suspended between porcelain supports. Fig. 3(B) gives the curves of probe wire versus guard cylinder potential in the cylindrical geometry for various positions. It may be seen that the curves are flat in the proper region, indicating that the probe is operating properly. Fig. 4 shows a comparison of calculations and measured potentials, the circled points for a wire voltage of 23.5 kv being obtained from the curves of Fig. 3(B). The points in Fig. 4 shown in triangles were measured at a much earlier date (approximately one year earlier) than those shown in circles. If this can be used as a measure of the reproducibility of the measurements, then one would say the results are quite reproducible. Again, it is pointed out that all potentials are actually negative. It may be seen that calculated and measured values are in good agreement, the maximum error being approximately 10%. The measured points are consistently higher than the calculated curves which can be accounted for, at least in part, by the error potential previously discussed.

Potential Measurements for Wires Between Parallel Plates

Wires between parallel plates are widely used for industrial precipitators but the resulting potentials have not been amenable to calculation. The probe method is useful for measuring potentials in this situation. Measurements have been made for various conditions in the geometry consisting of wires spaced 6 inches from center to center midway between parallel grounded planes spaced 9 inches apart. This arrangement is shown in Fig. 5. Figs. 6(A) and (B) show measurements of space potentials for wire voltages of 25.5 and 43.5 kv respectively. These measurements are precise and re-

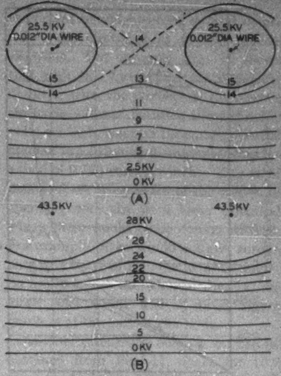


Fig. 8. Equipotential lines. A—As obtained from Fig. 6(A). B—As obtained from Fig. 6(B)

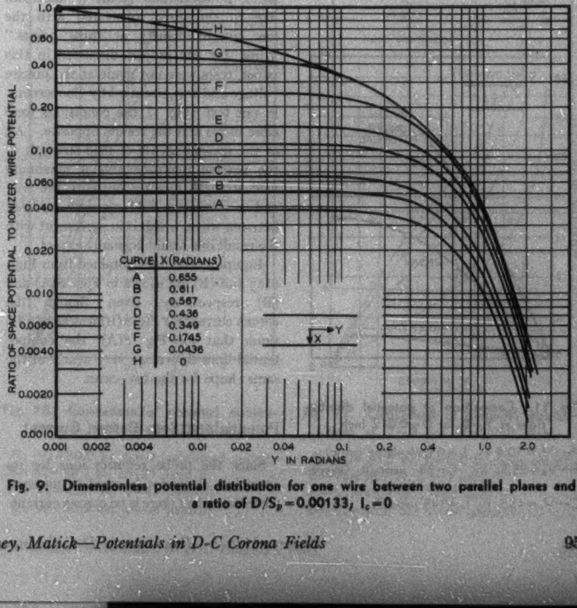


Fig. 9. Dimensionless potential distribution for one wire between two parallel planes and a ratio of $D/S_p = 0.00133$, $I_c = 0$

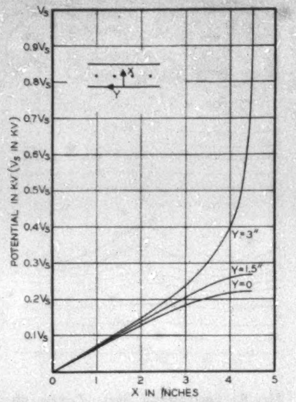


Fig. 10. Normalized potential distribution at zero corona current for four wires of 0.012-inch diameter between two infinite parallel planes

peatable, and are believed to be quite accurate.

Potentials are measured to within 1.5 inches of the wire. The remainder of the curve is shown as a dashed line. As a basis for drawing the upper portion of the dashed curve, the slope at the wire was calculated with the use of the following equation:

$$E_0 = \frac{V_0}{r_0 \ln \dfrac{r_0}{R_0}} \qquad (2)$$

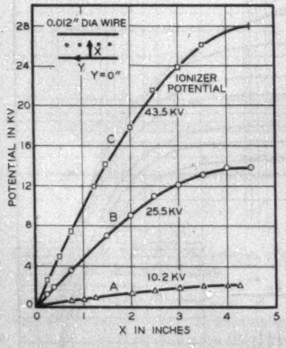


Fig. 11. Comparison of potential distributions at Y=0 for D=0.012 inch

A—V_s=10.2 kv, I_c=0 (calculated curve)
B—V_s=25.5 kv, I_c=35 μamp/ft (microamperes per foot)
C—V_s=43.5 kv, I_c=145 μamp/ft

where

E_0 = electric field intensity at wire surface
V_0 = voltage necessary to initiate corona
r_0 = radius of wire
R_0 = distance to passive electrode

This is the equation for a wire in a concentric cylinder. At the wire the field is influenced relatively little by the shape of the passive electrode and so this should be a good approximation to the field at the wire for a wire between parallel planes. This equation, of course, assumes that the field at the wire is independent of corona current so long as corona exists.

The condition at the passive electrode is of primary importance in determining the rate of precipitation. To study the condition in this region, potentials were also measured along lines parallel to the passive electrodes. These measurements are plotted in Fig. 7.

Experimental results obtained by Cooperman[1] and shown in his Fig. 4 give a field strength near the passive electrode of 17.6 electrostatic units directly opposite a wire, as compared with 27 midway between wires. These figures indicate that the electric field strength is more than 50% greater between wires than directly opposite a wire. This is directly opposite to the result which one would normally expect and so started considerable discussion. Cooperman says, "The technique was not pushed to a high level of accuracy . . ."; in his conclusion he says, "There are indications that the field-strength at the plate surface is maximum between the wires, however this result is uncertain due to the error associated with the low current density at these points." While, for the project discussed in this paper, tests were not made at the precise voltage and wire size used by Cooperman, in one test (Fig. 7) the conditions were close to his and the results opposite. All of the measurements of field-strength at the passive electrode give a maximum field strength directly opposite the wire and a minimum field strength midway between wires, which is in agreement with the result one would normally expect.

Equipotential lines obtained from Figs. 6(A) and (B) are shown in Figs. 8(A) and (B) respectively. Even though the corona current for Fig. 8(B) was nearly 3.5 times that for Fig. 8(A), the equipotential lines shown are very nearly of the same shape for the two cases.

Potentials at Zero Corona Current

Since the probe requires ions for its operation, it cannot be used to measure potentials when there is no corona current.

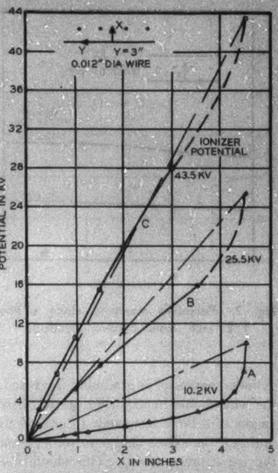


Fig. 12. Comparison of potential distributions at Y=3 for D=0.012 inch

A—V_s=10.2 kv, I_c=0 (calculated curve)
B—V_s=25.5 kv, I=35 μamp/ft
C—V_s=43.5 kv, I_c=145 μamp/ft

However, in this case, the equation governing the space potential is linear, and so potentials can be calculated for many boundary conditions. For a wire between parallel planes, the potential can be calculated using the method of conformal transformation. The method is discussed in the Appendix. Dimensionless curves showing the potentials in space for unit potential on a wire with $D/S_p=0.00133$ are given in Fig. 9. As explained in the Appendix, potentials for other wire sizes can be obtained from this same set of curves provided that D/S_p is not less than 0.00133. Since the problem is linear, potentials for more than one wire can be obtained by the super-

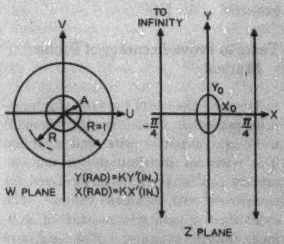


Fig. 13. Equivalent geometries resulting from transformation w=tan Z

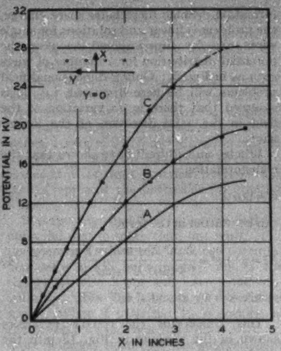


Fig. 14. Comparison of potential distributions at $Y=0$ for a wire potential of 43.5 kv

A—$D=0.160$ inch, $I_o=0$ (calculated curve)
B—$D=0.080$ inch, $I_o=48$ μamp/ft
C—$D=0.012$ inch, $I_o=145$ μamp/ft

position of potentials for the individual wires. In this way, the potential distribution for the four wires illustrated in Fig. 5 were calculated. These potentials for four wires, each 0.012 inch in diameter are shown in Fig. 10. Potentials for two other wire sizes are given by curves *A* in Figs. 14, 15, 16 and 17.

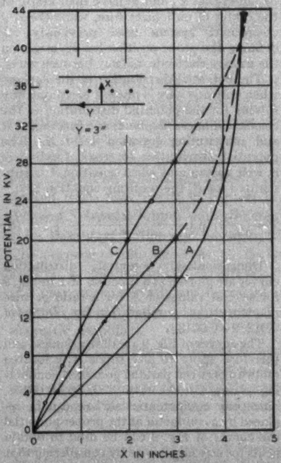


Fig. 15. Comparison of potential distributions at $Y=3$ for a wire potential of 43.5 kv

A—$D=0.160$ inch, $I_o=0$ (calculated curve)
B—$D=0.080$ inch, $I_o=48$ μamp/ft
C—$D=0.012$ inch, $I_o=145$ μamp/ft

Effect of Corona Current on Potential Distribution

It is very desirable to know how the potential distribution and hence electric field intensity depend on the corona current. The corona current can be varied by changing the wire potential with the electrodes unchanged, or it can be varied by changing the wire size while keeping the potential constant. The curves of Figs. 14 and 15 show space potentials when 43.5 kv is applied to wires ranging from 0.012- to 0.16-inch diameter. The curves of Figs. 11 and 12 show how the field distribution changes with corona current for wires 0.012 inch in diameter. The curves of Figs. 16 and 17 show a similar comparison for wires 0.080-inch diameter. The slope of the curves in Figs. 11, 12, and 14 through 17 gives directly the electric field intensity. The reason is that because of symmetry, there exists only an x component of electric field intensity along the lines $y=0$ and 3 inches. Furthermore, because of symmetry, the total electric field intensity evaluated at the point $x=4.5$, $y=0$ inches must be identically equal to zero. It may be seen that all the results are in agreement with this. These curves illustrate the extent to which corona current determines the electric field at the passive electrode. For example Fig. 14 shows that with 43.5 kv applied to a 0.16-inch-diameter wire, the gradient at the passive electrode midway between wires is 4,250 volts per inch, whereas the same potential applied to 0.012-inch-diameter wires gives a field of 9,750 volts per inch. This is, of course, directly opposite to the results of changing wire size at a lower potential where there would be no corona current. For example, 10 kv would not give corona on a wire 0.012 of inch diameter or larger. So for 10 kv on the wires of Fig. 14, the larger wire would give a much higher gradient at the passive electrode.

Another interesting comparison can be obtained from Fig. 11 where for 10.2 kv applied to wires 0.012 inch in diameter the gradient at the passive electrode and midway between wires is 750 volts per inch, but if the potential is increased to 43.5 kv, the gradient is 10,000 volts per inch. Thus, an increase of 327% in potential on the wires gives an increase in gradient at the passive electrode of 1,123%.

All these measurements have been made for the case of no dust loading. Heavy dust loadings will have some effect

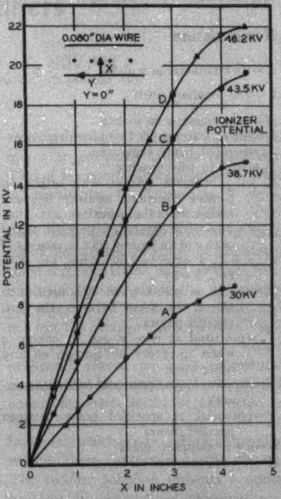


Fig. 16. Comparison of potential distributions at $Y=0$ for $D=0.080$ inch

A—$V_s=30.0$ kv, $I_o=0$ (calculated curve)
B—$V_s=38.7$ kv, $I_o=21$ μamp/ft
C—$V_s=43.5$ kv, $I_o=45$ μamp/ft
D—$V_s=46.2$ kv, $I_o=64$ μamp/ft

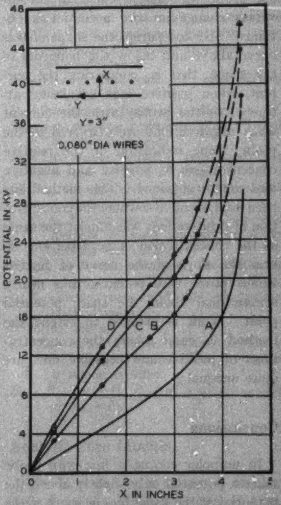


Fig. 17. Comparison of potential distributions at $Y=3$ for $D=0.080$ inch

A—$V_s=30.0$ kv, $I_o=0$ (calculated curve)
B—$V_s=38.7$ kv, $I_o=21$ μamp/ft
C—$V_s=43.5$ kv, $I_o=45$ μamp/ft
D—$V_s=46.2$ kv, $I_o=64$ μamp/ft

on the space potential distribution but such cases have not been studied.

Measurement of Abnormal Corona

In the case of the charged-ball method for measuring a field any ions of the wrong sign destroy the validity of the method since even a very small percentage of ions of the wrong sign gives a very large reduction in the equilibrium charge acquired by the ball. However, in the probe method which has been described here, any ions of the wrong sign cause the probe to assume more accurately the potential of the space to be measured and thus reduce the error potential. If there could be equal numbers of positive and negative ions and their mobilities were equal, there would be no error potential due to excess charge. In contrast, the ball-dropping method would give no charge and therefore no measurement at all.

With abnormal corona, in addition to measuring potential, the probe can be used to obtain some information regarding the concentration of ions of each sign. The probe frame or guard is held at the potential of the space as determined in the usual manner. Then, with a high-impedance circuit a current is drawn from the probe wire and a curve of current versus change in wire potential is obtained. By comparing the shape of this curve above and below the potential of the space, that is, comparing the current, when positive ions are being attracted to the probe, with the current when negative ions are attracted to the probe, some indication of the relative concentrations of positive and negative ions can be obtained. This method has been used in one case where the concentration of positive ions was almost the same as the concentration of negative ions so that the slope of the curve of current versus probe wire potential was nearly symmetrical about the space potential point. Work is planned to extend the method to cases where the concentrations of positive and negative ions are quite unequal.

Conclusions

In unipolar corona a fine wire probe takes a potential only slightly above the potential of the same line in space if the probe were removed. The method of gradually increasing the potential of the frame and guard, and recording wire potential as a function of guard potential gives a region over which the guard potential can be varied without changing the wire potential. This interval of unvarying wire potential is the true probe potential or is the potential of the space plus the small error potential; usually this error potential is negligible. For more precise work, since the error potential is small, a reasonable approximation of the error potential will bring the result within the region of instrument uncertainties.

With this method, potentials can be measured quite precisely throughout the corona region. The field close to the large or passive electrode of a precipitator very largely determines the rate of precipitation and so is of primary importance. The probe method permits measuring this field with good accuracy. For the cases investigated, the charge associated with the corona current has very little effect on the direction of the equipotential line at any point in space. But there is a large change of gradient of potential with corona current, so that for a wire potential of the order of 50 kv the gradient at the passive electrode is greater for fine wires than for large wires.

The probe method can be used for measurements in abnormal corona where ions of both signs exists. In this case the error potential is less than it is where there are ions of only one sign.

Nomenclature

A = wire diameter in w plane
$k = \frac{\pi}{4} \times \frac{S_p}{2}$ radians/inch
D = wire diameter in inches
I_1 = corona current per foot of ionizer wire
S_p = parallel plane separation
S_w = wire to wire spacing
x, y = co-ordinate system with origin on passive electrode, midway between ionizer wires (in inches)
X, Y = co-ordinate system with origin at the center of an ionizer wire in radians
x', y' = X, Y expressed in inches, i.e., $X = kx'$, $Y = ky'$
q_n = charge on wire n when V_1 is applied to all of the wires between the two parallel planes
V_n = potential at wire n due to charge q when there are no charges on any other wires
V_s = applied ionizer potential (all wires present)
V = potential at specified point between parallel planes
$w = u + iv$ complex plane
$Z = X + iY$, complex plane

Appendix

When there is no space charge present (i.e., no corona current) it is possible to calculate the potential distribution for a single wire midway between two parallel planes by making use of a conformal transformation. Since there is no space charge, the problem is linear and solutions for single wires can be superimposed to give the potential distribution for a number of wires such as in Fig. 5. Only 2-dimensional field problems will be here discussed, i.e., it is assumed that there is no variation of the field in the axial direction of the ionizer wires.

Wholey and Eldred[6] have shown that the transformation

$$w = \tan Z \tag{3}$$

can be written in the form

$$R^2 = u^2 + v^2 = \frac{\tan^2 X + \tanh^2 Y}{1 + \tan^2 X \tanh^2 Y} \tag{4}$$

where $w = u + iv$ and $Z = X + iY$.

This transforms the coaxial cylinders shown in the w-plane in Fig. 13 into the elliptical wire midway between the two parallel planes shown in the Z-plane in Fig. 13. The two parallel planes in the Z-plane (i.e., $X = \pm\pi/4$) correspond to the outer (larger) cylinder in the w-plane (i.e., $R = 1$), while the elliptical wire in the Z-plane corresponds to the inner (small) cylinder in the w-plane (i.e., $R = A$).

In the Z-plane, X and Y are expressed in radians, while in the w-plane u and v have no dimensions. Furthermore, $\pi/4$ radians in the Z-plane must correspond to one half the parallel plane separation in the actual geometry. Let x' and y' be the co-ordinates in inches corresponding to X and Y (in radians) respectively. The solution in the Z-plane is general since the grounded plane separation can be changed merely by changing k, and the wire diameter can be changed by changing X_0 (which in turn determines Y_0, since $A = \tan X_0 = \tanh Y_0$). It should be noted that x', and y' have their origin at the center of an ionizer wire, while the x, y co-ordinate system used previously for defining the probe position has its origin on the passive electrode midway between wires.

The potential distribution in the Z-plane when only one wire is present is obtained by solving for the potential distribution in the coaxial cylinders (w-plane) as a function of R and substituting equation 4 for R. The potential distribution for coaxial cylinders is well known and when equation 4 is substituted for R, the resulting equation is

$$V = \frac{V_n}{2 \ln A} \ln \left(\frac{\tan^2 X + \tanh^2 Y}{1 + \tan^2 X \tanh^2 Y} \right) \tag{5}$$

Dimensionless potential distribution curves are plotted as a function of Y in Fig. 9 for several values of X for a ratio or wire diameter to plane separation (i.e., D/Sp) of $0.012/9 = 0.00133$.

The curves of Fig. 9 are dimensionless and therefore apply to any single wire located midway between parallel grounded conducting planes for the ratio. $D/Sp = 0.00133$. Since any equipotential surface can be replaced by a conductor at the proper potential the curves of Fig. 9 can be used to obtain fields for any larger wire by considering that the proper equipotential is replaced by a conducting suface (wire) at the appropriate potential. For example, for a wire of 0.02096 radian diameter (i.e., a 0.12-inch-diameter wire between two planes 9 inches apart) the curves of Fig. 9 give a wire potential of 0.66 at a radius of 0.06 inch

($X=0$, $Y=0.01048$ radians). The values as read from the curves are then the potential resulting from a potential of 0.66 on a 0.12-inch diameter.

Theoretically the equipotential surfaces obtained from Fig. 9 are ellipses and so theoretically the wires must be elliptical but, fortunately for small wires (for $Sp=9$ inches, a wire of 0.375-inch diameter can be considered small), the ellipses are so close to circles that for most practical purposes the curves represent the potentials for circular wires.

Superposition can then be used to find the potential due to more than one wire. For the geometry shown in Fig. 5, the potential near wire 1 will be primarily due to the charge on wire 1 but will have some contribution from the charges on the other wires. When a potential V_s is applied to the several wires, each individual wire acquires charges such that the superimposed fields give the potential V_s on each wire.

From Fig. 9, a potential of 1.00 on a wire of 0.012-inch diameter and $S_p=9$ inches produces a potential of 0.037 at $x'=0$, $y'=6$ inches ($X=0$, $Y=1.0408$ radians). At $x'=0$, $y'=12$ inches ($X=0$, $Y=2.096$ radians), the potential is only 0.0047. At 18 inches from the wire, the potential is less than 0.001 and is negligible. Thus at wire one

$$V_s=V_1+0.037V_2+0.0047V_3 \tag{6}$$

and at wire 2

$$V_s=V_2+0.037V_1+0.037V_3+0.0047V_4 \tag{7}$$

By symmetry

$$V_1=V_4 \text{ and } V_2=V_3$$

Substituting these values into the above equations and solving yields

$$V_1=0.96V_s \text{ and } V_2=0.925V_s$$

If this method is extended to wires which are larger and closer together, one must be sure that the ellipse representing the wire is sufficiently close to the actual shape and that when potentials are superimposed, the resulting potential is sufficiently uniform over the cross section of any wire.

References

1. A New Technique for the Measurement of Corona Field Strength and Current Density in Electrical Precipitation. *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 75, Mar. 1956, pp. 64–66.

2. Some Measurements of Abnormal Corona, G. W. Penney, *Ibid.*, vol. 77, July 1958, pp. 319–27.

3. Particle Charging in Electrostatic Precipitation, H. J. White. *Ibid.*, vol. 70, pt. II, 1951, pp. 1186–91.

4. Potentials in DC Corona Fields, R. E. Matick. *Doctoral Dissertation*, Carnegie Institute of Technology, Pittsburgh, Pa., chap. I.

5. *Ibid.*, chap. III.

6. A New Type of Slotted Line Section, W. B. Wholey, W. N. Eldred. *Proceedings*, Institute of Radio Engineers, New York, N.Y., vol. 38, Mar. 1950, pp. 244–48.

7. Conduction of Electricity Through Gases, J. J. Thomson. Cambridge University Press, Cambridge, England, vol. 2, third edition, 1931, p. 552.

8. Potentials in DC Corona Fields, R. E. Matick. *Doctoral Dissertion*, Carnegie Institute of Technology, chap. V.

Rectifier Unblocking From a Generalized Analysis of Self-Saturating Magnetic Amplifiers

H. C. BOURNE
MEMBER AIEE

D. NITZAN
ASSOCIATE MEMBER AIEE

THE UNBLOCKING of the self-saturating gate rectifier of the resetting core in magnetic-amplifier circuits is a well-known phenomenon. Circuit behavior during this unblocked period, a condition sometimes called gate-limited reset, may result in appreciably different response times and different gains than would otherwise be expected. The purpose of this paper is to derive analytic predictions of the occurrence of this phenomenon which will be useful in the analysis and design of self-saturating magnetic amplifiers.

The analytic prediction of gate-limited reset will be obtained from a generalized analysis applicable to the three standard single-phase self-saturating magnetic-amplifier circuits. Such an analysis requires approximate analytic representations of the nonlinear solid-state elements in the circuits, the ferromagnetic cores, and the semiconductor rectifiers. Further, since phenomena associated with instantaneous values are to be examined, these representations must give approximate information about the instantaneous state of the nonlinear elements and not just the average characteristics. Another severe restriction, which is rather arbitrarily placed on these representations, is that the resulting formulation of circuit behavior should be amenable to reasonable analytic treatment. Two core functions, which are as realistic as possible within these restrictions, are chosen. The rectifier function presents a minor problem in comparison. Experimental results will give an indication of the validity of the representations and the degree of approximation involved.

The combination of an alternating square-wave gate supply and d-c control, and the combination of sinusoidal gate supply and rectified sinusoidal control, are examined.

Nomenclature

e_s, E_s = instantaneous and half-cycle average gate supply voltage
R_L, R_0 = load and total series gate resistance
I_{md}, I_{m0} = core-magnetizing currents corresponding to full and zero reset
G = core conductance
R_s = internal resistance of gate supply voltage
R_f = forward resistance of rectifier
R_w = gate-winding resistance plus forward rectifier resistance
$R_x=1/G_x$ = gate winding resistance plus reverse rectifier resistance
N_c, N_G, N = control turns, gate turns, and turns ratio N_G/N_c
e_1, E_1, e_2, E_2 = instantaneous and half-cycle average induced voltages in a resetting and gating core
eL, EL, e_0, E_0 = instantaneous and half-cycle average net and gross output voltage
e_c, E_c = instantaneous and half-cycle average control voltage
α = firing angle
e_{res}, E_u = instantaneous voltage across a rectifier and forward bias voltage of a silicon rectifier
$E_{c,u}$, $R_{c,u}$, α_u = E_c, R_c, and α corresponding to $e_{res}=E_u$

Analytic Core Models

The loop and node equations of magnetic-amplifier circuits contain terms which involve the induced voltage and the magnetizing current associated with ferromagnetic cores. Unfortunately, the relationship between these two variables, which describe the terminal properties of the core winding, depends on the present

---

Paper 60-91, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., January 31–February 5, 1960. Manuscript submitted October 21, 1959; made available for printing November 25, 1959.

H. C. Bourne is with the University of California, Berkeley, Calif., and D. Nitzan is with the Stanford Research Institute, Menlo Park, Calif.

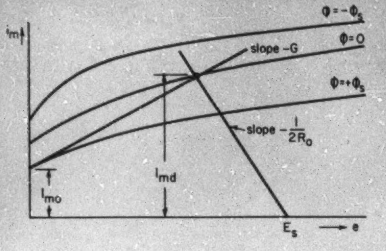


Fig. 1. Graphical representation of I_{md}, I_{m0} and G on core characteristics (magnetizing current versus excitation voltage with flux as a parameter. Initial flux corresponds to saturation)

and past states of the core material. Even if a common initial state, e.g., positive saturation, is assumed, the relationship between net magnetomotive forces F and the rate of change of flux $d\phi/dt$, with flux ϕ as a parameter, is nonlinear.[1] In a previous paper, a test procedure is developed by which the nonlinear characteristics of a resetting core may be obtained.[2] A graphical procedure is then devised which accurately predicts the core behavior in the prefiring and postfiring intervals regardless of the associated circuitry, e.g., the control circuit resistance.

In an analytic treatment these nonlinear core characteristics must be sufficiently linearized to give realistic results with reasonable effort. Two core models, designated A and B are considered. Core model A approximates the magnetizing current of a square-loop material by I_{md} during the prefiring period (mode 1) and by I_{m0} during the postfiring period (mode 2). Both symbols imply referral to the gate circuit. I_{md} is defined as the average magnetizing current when the core is fully reset by $E_s - 2I_{md}R_0$. I_{m0} is by definition the threshold magnetizing current corresponding to the static **B-H** loop at saturation. I_{md} and I_{m0} are related to typical core characteristics as shown in Fig. 1.

Approximate values of I_{md} and I_{m0} may be obtained by other means. To a reasonable approximation, I_{md} may be identified with the dynamic coercive force under rated sinusoidal excitation. Alternatively, I_{md} and I_{m0} may be identified with the magnetizing current at full and zero reset in the constant-current reset test. For small values of R_c a smaller value for I_{m0} than the one defined by Fig. 1 should be taken as a result of initial gating along minor loops before the end of the postfiring period.

Core model B involves a mathematically continuous rather than discontinuous linear approximation of the core characteristics and is inherently a more satisfying representation. After Kikuchi[3] and Feth[1] the magnetizing current is given by

$$i_m = I_{m0} + Ge \tag{1}$$

I_{m0} is the same as defined previously, and the core conductance G may be evaluated as

$$G = \frac{I_{md} - I_{m0}}{E_s - 2I_{md}R_0} \tag{2}$$

so that I_{md} corresponds to the magnetizing current at full reset in both cases; see Fig. 1. Neither core function predicts the changes in gain or the shearing of the transfer characteristic caused by different values of control circuit resistance. As the analysis will show, both core functions will inherently linearize the predicted operation from zero to full output.

Circuit Equations

Since phenomena associated with the instantaneous states of the nonlinear magnetic amplifier circuits are to be examined, equations relating instantaneous voltages and currents to the known excitations must be derived. The circuit diagrams of the doubler, bridge, and center-tap circuits are shown in Fig. 2. In the normal range of contol during the prefiring interval of the positive half-cycle (mode 1) core 1 is resetting from saturation and core 2 is gating. The corresponding equivalent circuits are shown in Fig. 3, with the leakage current of the blocked complementary rectifier of the bridge circuit neglected. During the postfiring interval of the positive half-cycle (mode 2), core 2 has saturated and the equivalent circuits are still valid

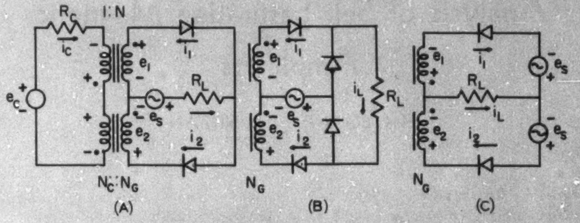


Fig. 2. Self-saturating magnetic-amplifier circuits; control circuits identical

A—Doubler gate and control circuits
B—Bridge gate circuit
C—Center-tap gate circuit

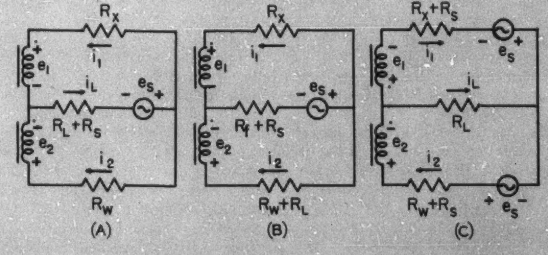


Fig. 3. Equivalent gate circuits of self-saturating magnetic amplifiers during positive half-cycle of gate excitation

A—Doubler
B—Bridge
C—Center-tap

Table I. Self-Resistances and Mutual Resistances of Gate Loops

	Doubler	Bridge	Center Tap
R_{11}	$R_x+R_L+R_s$	$R_x+R_f+R_s$	$R_x+R_L+R_s$
$R_{22}(=R_0)$	$R_w+R_L+R_s$	$R_w+R_f+R_s+R_L$	$R_w+R_L+R_s$
R_{12}	R_L+R_s	R_f+R_s	$-R_L$

with $e_2=0$, and the magnetizing current constraint removed from core 2.

The three loop equations and the two core-constraint equations based on core model *B* and applicable during mode 1 are

$$\begin{Vmatrix} 0 & 0 & R_c & 1/N & 1/N \\ R_{11} & R_{12} & 0 & 1 & 0 \\ R_{12} & R_{22} & 0 & 0 & 1 \\ 1 & 0 & -1/N & -G & 0 \\ 0 & 1 & 1/N & 0 & -G \end{Vmatrix} \begin{Vmatrix} i_1 \\ i_2 \\ i_c \\ e_1 \\ e_2 \end{Vmatrix} = \begin{Vmatrix} e_c \\ e_s \\ e_s \\ I_{m0} \\ I_{m0} \end{Vmatrix} \tag{3}$$

in which R_{11}, R_{22} $(=R_0)$, and R_{12} are the self-resistances and mutual resistances of the gate loops; see Table I. Equation 3 may be modified for core model *A* by substituting $G=0$ and I_{md} for I_{m0}. During mode 2 the equations apply if the last row and last column of the matrix are eliminated, and they apply to core model *A* with $G=0$. Equations applicable to the negative half-cycle may be easily inferred from these by symmetry considerations although, by the same token, they would contain no new information.

The solution of these equations is straightforward even if somewhat tedious; it is necessary in order to predict approximately the occurrence of any instantaneous misbehavior in the circuit. Appropriate solutions of these equations will be used in the section dealing with gate-rectifier unblocking but, first, the dynamic operation of the circuits will be derived by a method not requiring such solutions.

Output Difference Equations

If instantaneous values (waveshapes and firing angle) are not of concern, then the equation describing the dynamic operation of the magnetic amplifiers in normal operation may be obtained directly. Alternating square-wave gate supply and d-c control is assumed initially.

CORE MODEL B

In core model B the control circuit loop equation is

$$Ne_c=N^2R_c\frac{i_c}{N}+e_2-e_1 \tag{4}$$

The magnetizing current constraint of the resetting core is

$$G_x\left(e_s-e_1-\frac{e_LR_{12}}{R_L}\right)-\frac{i_c}{N}=I_{m0}+Ge_1 \tag{5}$$

in which the first term represents the reverse current of the gate rectifier. Substitution of i_c/N from equation 5 into equation 4 and integration over the half-cycle, designated the *n*th half-cycle, yields

$$\begin{aligned} NE_c(n)=N^2R_c(G_xE_s-I_{m0})-\\ (1+GN^2R_c+G_xN^2R_c)E_1(n)+E_2(n)-\\ E_L(n)G_xN^2R_c\frac{R_{12}}{R_L} \end{aligned} \tag{6}$$

The half-cycle average voltages across the cores are related to the load voltages on this and the next half-cycle $(n+1)$ as follows

$$E_1(n)=E_1(n+1)=E_s-\frac{E_L(n+1)R_0}{R_L}$$

$$E_2(n)=E_s-\frac{E_L(n)R_0}{R_L} \tag{7}$$

Substitution of equations 7 into equation 6 results in the output difference equation:

$$\begin{aligned} E_L(n+1)=\frac{NR_L}{R_0}\frac{E_c(n)}{1+N^2R_c(G+G_x)}+\\ \frac{E_L(n)\left(1+G_xN^2R_c\frac{R_{12}}{R_0}\right)}{1+N^2R_c(G+G_x)}+\\ \frac{R_L}{R_0}\frac{N^2R_c(I_{m0}+GE_s)}{1+N^2R_c(G+G_x)} \end{aligned} \tag{8}$$

Dynamic transfer functions to any degree of approximation desired may be obtained directly from this linear first-order difference equation. The quantities often of chief interest are the voltage gain and the time constant associated with the exponential portion of the response, which is characterized also by a half-cycle delay. The gain is obtained by assuming steady-state conditions and finding E_L/E_c which gives

$$\begin{aligned} K_v=\frac{NR_L}{R_0}\frac{1}{N^2R_c\left[G+G_x\left(1-\frac{R_{12}}{R_0}\right)\right]}\\ =\frac{NR_L}{R_0}k_{v0} \end{aligned} \tag{9}$$

in which k_{v0} is the gross voltage gain on an equal turns basis. The time constant is given by

$$\tau=\frac{1}{\ln\left[\frac{1+N^2R_c(G+G_x)}{1+\frac{G_xN^2R_cR_{12}}{R_0}}\right]}\cong\frac{1}{\ln\frac{k_{v0}+1}{k_{v0}}} \tag{10}$$

half cycles. For time constants greater than 4 or 5 half-cycles, the natural logarithm may be approximated by the first two turns of its series expansion to give

$$\tau\cong\frac{1}{N^2R_c\left[G+G_x\left(1-\frac{R_{12}}{R_0}\right)\right]}=k_{v0} \tag{11}$$

CORE MODEL A

The same method applied to core model *A* is somewhat more complicated because of the discontinuous nature of the core function. Integration of the control circuit loop equation over a half-cycle gives the equivalent of equation 6 as

$$\begin{aligned} NE_c(n)=N^2R_c\left[G_xE_s-\frac{I_{md}\alpha(n)}{\pi}-\right.\\ \left.I_{m0}\left(1-\frac{\alpha(n)}{\pi}\right)\right]-(1+G_xN^2R_c)E_L(n)+\\ E_2(n)-\frac{E_L(n)G_xN^2R_cR_{12}}{R_L} \end{aligned} \tag{12}$$

Equation 7 still applies, and in addition:

$$\begin{aligned} \frac{\alpha(n)}{\pi}\cong\frac{E_2(n)}{(E_s-2I_{md}R_0)}\\ =\frac{\left(E_s-\frac{E_L(n)R_0}{R_L}\right)}{E_s-2I_{md}R_0} \end{aligned} \tag{13}$$

Substitution of equations 7 and 13 into equation 12 gives the output difference equation for this core model.

$$\begin{aligned} E_L(n+1)=\frac{NR_L}{R_0}\frac{E_c(n)}{1+N^2R_cG_x}+\\ \frac{1-N^2R_c\left(G-\frac{G_xR_{12}}{R_0}\right)}{1+N^2R_cG_x}E_L(n)+\\ \frac{R_L}{R_0}\frac{N^2R_c(I_{m0}+GE_s)}{1+N^2R_cG_x} \end{aligned} \tag{14}$$

in which $G=(I_{md}-I_{m0})/(E_s-2I_{md}R_0)$ has been defined exactly as in core model *B*.

The voltage gain from equation 14 is identical with that obtained for core model *B* and, in fact, the steady-state characteristics predicted by these core models are identical in every respect. This is not surprising since both core models result in a linear transfer characteristic between the same end points. In theory, the time constant differs, since, from this core model

$$\tau=\frac{1}{\ln\left[\frac{1+N^2R_cG_x}{1-N^2R_c\left(G-\frac{G_xR_{12}}{R_0}\right)}\right]}\cong\frac{1}{\ln\frac{k_{v0}}{k_{v0}-1}}\ \text{half cycles} \tag{15}$$

However, for time constants greater than

4 or 5 half-cycles the appropriate approximation predicts the same result as obtained for core model *B* in equation 11. Actually in practical cases, for the range of time constants less than 4 or 5 half-cycles, the unblocking of the gate rectifier of the resetting core invalidates both "exact" expressions given by equations 10 and 15, as will be shown. Experimental results with the rectifiers replaced by switching transistors in order to force operation in the normal modal sequence indicate that equation 10 gives better correlation for fast response times.

In all the above equations if rectifier reverse current is small compared with magnetizing currents ($G>>G_x$), G_x may be made zero to obtain the corresponding, less complex, expressions.

Sinusoidal Excitation

In the previous derivation of the difference equation, alternating square-wave gate excitation and d-c control have been assumed. These results may be extended to sinusoidal gate excitation and rectified sinusoidal control with a few additional approximations.

Both core models bear considerably less resemblance to the actual core function if sinusoidal excitation is used rather than square-wave excitation. Core model *A* is particularly poor in mode 2 with low control resistance. Core model *B*, if applied to instantaneous values, as was intended, predicts a sinusoidally bulging **B-H** loop, which is certainly not observed experimentally. However, the following approximations are suggested.

In the case of core model *A* with sinusoidal excitation, the transcendental modes at the beginning and end of the half-cycle, which exist because of the theoretical inability of the excitations to supply I_{md} and I_{m0} instantaneously, are neglected. Further, the sinusoidal variation of the half-cycle-average output voltage with firing angle is approximated by the linear function which agrees at $\alpha=0$ and $\alpha=\pi$. With these two approximations the results are identical with those for square-wave excitation.

In the case of core model *B*, if the transcendental modes are neglected as before, and the **B-H** loop is assumed to have the sinusoidally bulging sides predicted by this core model, then the results are obviously the same as those previously obtained. No approximation regarding the firing angle versus output is necessary. However, if equation 1, defining this model, is assumed to apply to half-cycle average values, the results are again the same as those obtained previously. Of course, in this case the assumption concerning the linear relationship between firing angle and output is necessary.

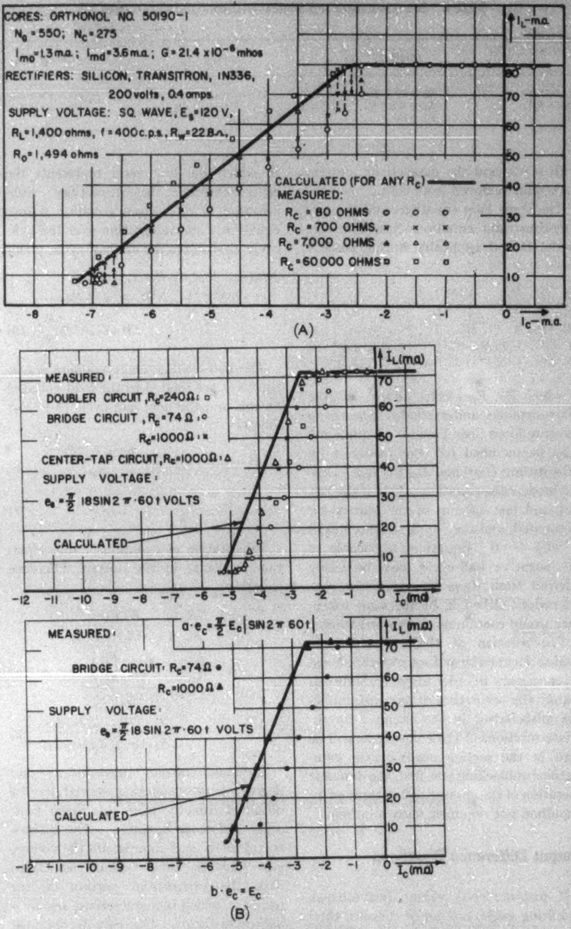


Fig. 4. Steady-state transfer characteristics with normal rectifier behavior

A—With square-wave excitation
B—With sinusoidal excitation

With approximations comparable to those suggested above, equations 8 and 14 are also applicable to sinusoidal gate excitation and d-c control.

Experimental Verification

Rectifier unblocking, when present, is eliminated experimentally by the use of switching transistors which are synchronized with the supply voltage.

The circuitry common to square-wave and sinusoidal excitation is as follows: Cores, Magnetics Inc.; Orthonol no. *50190;* 1-mil tape; inside diameter = 1 inch; outside diameter = 1.75 inch; height = 0.5 inch; $l_{Fe}=10.96$ centimeters; $A_c=0.909$ square centimeter; $G=21.4\times10^{-6}$ mhos, $N_G=550$; $N_c=275$; $R_f=4.6$ ohms; $R_w=22.8$ ohms; $R_x\cong\infty$ (silicon recti-

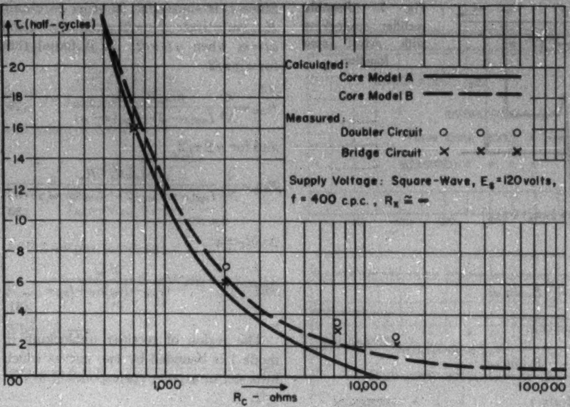


Fig. 5. Time constant versus R_c

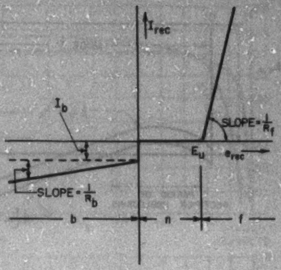


Fig. 6. Semiconductor rectifier model

fiers, transitron *1N336;* 200 volts; 0.4 ampere).

SQUARE-WAVE EXCITATION

$E_s = 120$ volts, $f = 400$ cycles per second, $R_L = 1{,}400$ ohms, $R_s = 71$ ohms, $I_{md} = 3.6$ ma (milliamperes), $I_{m0} = 1.3$ ma. Fig. 4(A) shows a comparison between calculated and experimental transfer characteristics.

The unstable region of ferroresonance nature around full output when R_c is small is caused by the winding capacitance of the resetting core.

An output voltage transient response to a step control voltage is half-cycle integrated and recorded by a Polaroid camera. The resulting time constant values for various values of R_c are compared in Fig. 5, with the calculated values based on core models *A* and *B*. For a control resistance $R_c = 1/N^2G$, in this case 11,700 ohms, core model *A* predicts half-cycle response. For $1/N^2G < R_c < 2/N^2G$ core model *A* predicts stable decaying oscillations, for $R_c = 2/N^2G$ sustained oscillation at a frequency half of the supply frequency and for $R_c > 2/N^2G$ unstable exponential oscillations. Of course none of these predictions are found experimentally. The breakdown of this core model at large R_c stems from the assumption that $I_{md} - I_{m0}$ stays constant regardless of R_c. Actually $I_{md} - I_{m0}$ decreases as R. increases and, when $R_c \to \infty$ $I_{md} - I_{m0} \to 0$. Core model *B* predicts closer results, especially when R_c is large. Its predictions of infinite R_c for half-cycle response is verified experimentally. For low values of R_c the two predicted curves approach each other and have better correlation with measured values.

Table II. Calculated and Measured Time Constant with Sinusoidal Excitation

Circuits	Rc, Ohms	Calculated τ, Half-Cycle	Measured τ, Half-Cycle Rising E_L	Measured τ, Half-Cycle Falling E_L
Doubler	240	15	15	16
Bridge	500	7	8	9
Center tap	1,000	4	5	6

SINUSOIDAL EXCITATION

Here $e_s = \pi/2 \times 18 \sin 2\pi \times 60\, t$ volts, $R_L = 218$ ohms, $R_s = 2$ ohms, $I_{md} = 2.65$ ma, and $I_{m0} = 1.3$ ma. Experimental and calculated transfer characteristics and time constant are compared in Fig. 4(B) and Table II respectively.

The Criterion for Rectifier Unblocking

When $|e_c|$ is large, the reflected voltage across the gate side of the resetting core may unblock the corresponding self-saturating rectifier. For analysis purposes an amenable, yet realistic, rectifier model is desired. Such a model for a silicon rectifier is shown in Fig. 6.

Referring to Fig. 6, the rectifier of the resetting core is blocked when $e_{rec} \le E_u$. From Fig. 2 the blocking voltage across the rectifier of the resetting core, if the rectifier leakage current is neglected, is $e_1 + R_{12}i_L - e_s$. Consequently, the criterion for a blocked rectifier is

$$e_1 + R_{12}i_L - e_s \le E_u \tag{16}$$

The value of E_c, at which equation 16 becomes an equality, and referred to as $E_{c,u}$, may be found by substituting the expressions for e_1 and i_L from equation 3.

Rectifier Unblocking Under Square-Wave Excitation

MODE 1

Both core models *A* and *B* predict no rectifier unblocking in mode 1. This may be verified by showing that for $\alpha = \pi$, which is the extreme condition corresponding to e_c a minimum, $e_{rec} < 0$. However, with the actual instantaneous magnetizing current i_m, it is found that

$$E_{c,u} = -\frac{1}{N}\left[i_m(N^2R_c + 2hR_0) + 2E_u\right] \tag{17}$$

in which $h = 1 - R_{12}/R_0$, and rectifier unblocking is possible, as shown in Fig. 7.

MODE 2

In mode 2

$$i_L = \frac{E_s - E_u}{R_0} \tag{18}$$

Based on core model *A*

$$e_1 = -NE_c - I_{m0}N^2R_c \tag{19}$$

Substitution into equation 16 gives

$$E_{c,u} = -\frac{1}{N}\left[I_{m0}N^2R_c + hE_s + (2-h)E_u\right] \tag{20}$$

For a given α the corresponding control resistance is

$$R_{c,u} = \frac{1}{N^2}\,\frac{hE_s + (2-h)E_u}{\dfrac{\alpha}{\pi}(I_{md} - I_{m0})} \tag{21}$$

The maximum value of R_c that may be used without causing rectifier unblocking is determined with $\alpha = \pi$.

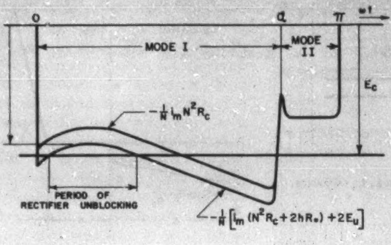

O
α
π
ωt
MODE I
MODE II
$E_{c,u}$
E_c
$-\frac{1}{N} i_m N^2 R_c$
PERIOD OF RECTIFIER UNBLOCKING
$-\frac{1}{N}\left[i_m(N^2R_c+2hR_0)+2E_u\right]$

Fig. 7. Possible rectifier unblocking with actual core function

$$R_{c,u(\pi)}=\frac{1}{N^2}\,\frac{hE_s+(2-h)E_u}{I_{md}-I_{m0}} \tag{22}$$

For $R_c\to\infty$, $\alpha=0$ and $I_{c,u(R_c=\infty)}=-NI_m{}^0$ Similarly, based on core model B

$$e_1=-\frac{NE_c+I_{m0}N^2R_c}{1+GN^2R_c} \tag{23}$$

$$E_{c,u}=-\frac{1}{N}\left\{I_{m0}N^2R_c+[hE_s+(2-h)E_u]\times(1+GN^2R_c)\right\} \tag{24}$$

$$R_{c,u,\pi}=-\frac{1}{N^2G}\,\frac{[hE_s+(2-h)E_u](1+2GR_0)}{E_s[1-h(1+2GR_0)]-E_u[1+(2-h)(1+2GR_0)]-2I_{m0}R_0} \tag{25}$$

and

$$I_{c,u(R_c=\infty)}=-N\{I_{m0}+G[hE_s+(2-h)E_u]\} \tag{26}$$

By substituting the values for h of each circuit and letting $\alpha=\pi$, it is concluded that according to core model A rectifier unblocking in mode 2 is possible in the three circuits, whereas core model B excludes this possibility from the center-tap circuit.

For design purposes, a region of rectifier unblocking is obtained by plotting $E_{L,u}$ (E_L corresponding to $E_{c,u}$) versus $E_{c,u}$ with R_c as a parameter. Evaluation of $E_{L,u}$ is based on the linear transfer characteristics assuming no rectifier unblocking. This step is valid since ($E_{c,u}$; $E_{L,u}$) is on the boundary between the regions of blocked and unblocked rectifier.

Rectifier Unblocking Under Sinusoidal Excitation

Rectifier unblocking under sinusoidal excitation becomes rather cumbersome due to the possible presence of transcendental equations depending on α. Additional complexity arises when rectifier unblocking occurs simultaneously in the two modes of operation.

Only core model A is considered. It is assumed that

$$e_s=\frac{\pi}{2}E_s\sin\omega t$$

and that

$$e_c=\frac{\pi}{2}E_c|\sin\omega t|$$

Mode 1

In mode 1, referring to Fig. 2,

$$e=e_s-E_u-Ne_c-I_{md}(N^2R_c+2R_0) \tag{27}$$

Substitution of equation 27 and $i_L=2I_{md}$ into equation 16 yields

$$E_{c,u}=-\frac{1}{N}\,\frac{2}{\pi\sin\omega t}(2hI_{md}R_0+2E_u+I_{md}N^2R_c) \tag{28}$$

Since e_c is sinusoidal, in order to obtain $E_{c,u}$, substitute $\omega t=\pi/2$ when $\alpha\geq\pi/2$ and $\omega t=\alpha$, when $\alpha\leq\pi/2$. It is found that for $\alpha\geq\pi/2$

$$R_{c,u}=\frac{4}{N^2}\,\frac{hI_{md}R_0+E_u}{I_{md}(\alpha-2)+I_{m0}(\pi-\alpha)} \tag{29}$$

and for $\alpha\leq\pi/2$

$$R_{c,u}=\frac{4}{N^2}\,\frac{hI_{md}R_0+E_u}{I_{md}(\alpha\sin\alpha-2)+I_{m0}\sin\alpha(\pi-\alpha)} \tag{30}$$

As $R_c\to\infty$

$$\alpha_{u(R_c=\infty)}=\sin^{-1}\frac{2I_{md}}{(I_{md}-I_{m0})\alpha_u+I_{m0}\pi} \tag{31}$$

The region of rectifier unblocking in mode 1 is bounded by two curves which intersect at a point corresponding to $\alpha=\pi/2$.

Mode 2

Following the same procedure, in mode 2

$$E_{c,u}=-\frac{1}{N}\left\{hE+\frac{2}{\pi\sin\omega t}\times[I_{m0}N^2R_c+(2-h)E_u]\right\} \tag{32}$$

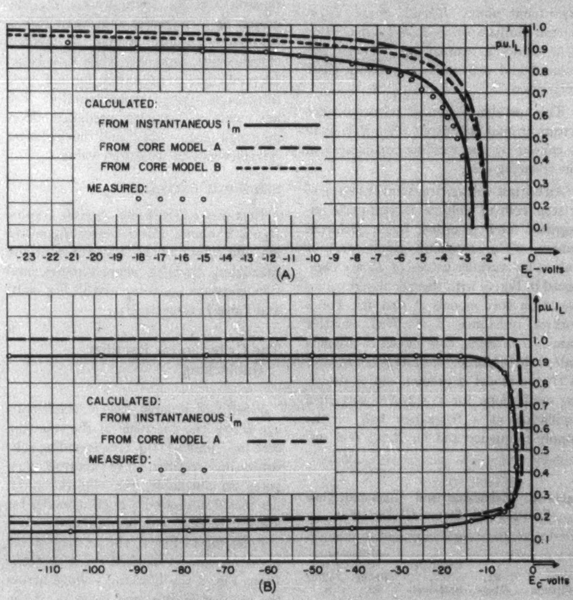


Fig. 8. Region of rectifier unblocking in Mode 2, doubler circuit

A—Square-wave excitation
B—Sinusoidal excitation

Contrary to mode 1, substitute $\omega t=\pi/2$, when $\alpha\leq\pi/2$ and $\omega t=\alpha$ when $\alpha\geq\pi/2$. When $\alpha\leq\pi/2$

$$R_{c,u}=\frac{1}{N^2}\frac{\pi hE_s+2(2-h)E_u}{I_{md}\alpha+I_{m0}(\pi-\alpha-2)} \qquad (33)$$

and when $\alpha\geq\pi/2$

$$R_{c,u}=\frac{1}{N^2}\frac{\pi hE_s+\dfrac{2(2-h)E_u}{\sin\alpha}}{I_{md}\alpha+I_{m0}\left(\pi-\alpha-\dfrac{2}{\sin\alpha}\right)} \qquad (34)$$

At full output, $\alpha=0$, so

$$R_{c,u(0)}=\frac{1}{N^2}\frac{\pi hE_s+2(2-h)E_u}{I_{m0}(\pi-2)} \qquad (35)$$

When $R_c\rightarrow\infty$

$$\alpha_{u(R_c=\infty)}=\sin^{-1}\frac{2I_{m0}}{(I_{md}-I_{m0})\alpha_u+I_{m0}\pi} \qquad (36)$$

MODES 1 AND 2

Rectifier unblocking may occur simultaneously in the two modes of operation. However, an unblocked current in one mode of operation has only a second-order effect on the unblocked current in the other mode of operation. As such, the equations derived previously for each mode separately may still be used to a good approximation. The over-all region of rectifier unblocking is the area covered by the region of rectifier unblocking in either mode of operation or in both modes.

Six combinations of rectifier unblocking are possible:

1. Complete in mode 1 alone.
2. Mode-limited in mode 1 alone.
3. Complete in mode 1 plus mode limited in mode 2.
4. Mode limited in modes 1 and 2.
5. Mode limited in mode 2 alone.
6. Complete in mode 2 alone.

Experimental Verifications

The circuit parameters are identical with those described in the experimental investigation with no rectifier unblocking. The regions of rectifier unblocking are found experimentally by decreasing E_c slowly for each value of R_c and recording $E_{c,u}$ and $E_{L,u}$ at the first appearance of the unblocked current. Calculated regions of rectifier unblocking based on the core models and an independently measured i_m, are compared with experimental regions in Figs. 8 and 9.

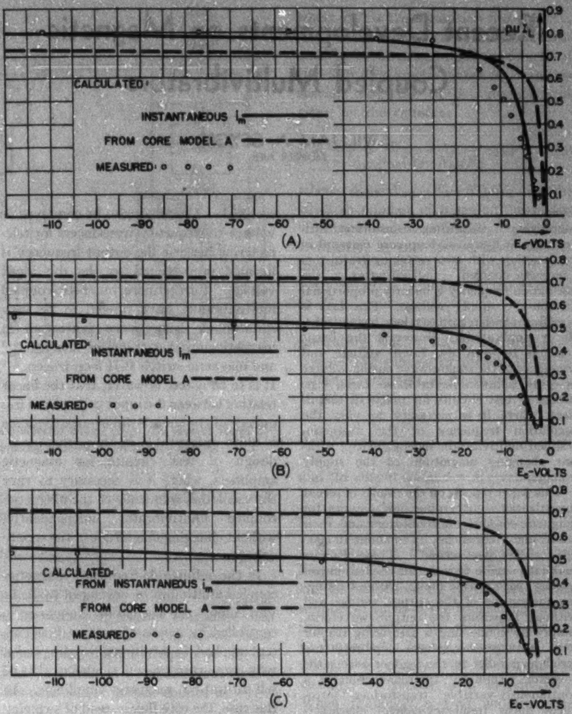


Fig. 9. Region of rectifier unblocking in mode I, sinusoidal excitation

A—Doubler
B—Bridge
C—Center tap

Conclusions

Two core models which yield linear relations may be used in a general analysis of the self-saturating magnetic amplifier. Difference equations, and steady-state and transient characteristics may be derived, assuming no rectifier unblocking. Based on this analysis, the criterion for rectifier unblocking may be determined. Regions of rectifier unblocking, superimposed on a family of steady-state transfer characteristics, may be calculated for various conditions and used to check analysis and design procedures.

The quantitative effect of rectifier unblocking on the time constant is discontinuous and, consequently, the concept of a time constant is no longer very meaningful. Theoretically, the increase in τ due to rectifier unblocking can be found by replacing R_t by R_u during the periods in which the rectifier is unblocked.

References

1. CORE-RESET FUNCTIONS IN MAGNETIC-AMPLIFIER ANALYSIS—PART I, G. C. Feth. *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 77, Sept. 1958, pp. 503–19.
2. GRAPHICAL EVALUATION OF MAGNETIC AMPLIFIER PERFORMANCE BASED ON CONSTANT-VOLTAGE RESET TEST, D. Nitzan. *Ibid.*, vol. 78, Nov. 1959, pp. 691–98.
3. ANALYSIS OF THE FULL-WAVE MAGNETIC AMPLIFIER CIRCUITS CONSIDERING THE CHANGE OF THE WIDTH OF THE DYNAMIC HYSTERESIS LOOP, Tadashi Kikuchi. *Ibid.*, vol. 76, July 1957, pp 241–49.

Recent Developments on Magnetic-Coupled Multivibrators

WILLIAM A. GEYGER
MEMBER AIEE

Synopsis: In a static d-c to a-c converter employing a saturating transformer with rectangular-hysteresis-loop core material in conjunction with two switching transistors (magnetic-coupled multivibrators), the oscillation frequency is linearly proportional to the d-c supply voltage. This fact is significant with regard to telemetering-transducer applications. However, the linear relation between d-c input voltage and frequency of the square-wave output voltage is quite unfavorable in those cases, e.g., design of test circuits for magnetic amplifiers, where it is necessary to vary the oscillation frequency of the magnetic-coupled multivibrator independently, without changing magnitude of the supply voltage.

This paper reports on the results of recent investigations concerning new circuits for magnetic-coupled multivibrators and their use at high frequencies, up to about 100,000 cps (cycles per second). These circuits make it possible to vary the oscillation frequency, within wide limits, without changing magnitude of the output voltage and without impairing the output waveform, simply by connecting a saturating reactor with rectangular-hysteresis-loop core material in parallel to the load or across the collector terminals. Interconnection of a continuously-variable transformer introduces linear frequency-control characteristics and permits coarse and fine adjustment with regard to frequency and magnitude of the square-wave output voltage.

A STATIC d-c to a-c converter employing a saturating transformer with rectangular-hysteresis-loop core material in conjunction with two switching transistors[1] is finding an increasing application as a square-wave power supply for magnetic amplifiers, induction motors, gyros, and torque relays. These magnetic-coupled multivibrators[2–5] are widely used in d-c power supply devices, where a low-voltage direct current is converted into a square-wave alternating current transformed to a high-voltage level and rectified. They may also be employed as voltage-to-frequency transducers for telemetering because the output frequency is linearly proportional to the d-c input voltage. Furthermore, magnetic-coupled multivibrators have proved to be very suitable as square-wave generators in magnetic-amplifier transient analyzers[6–8] and magnetic-switch B-H loop tracers.[9,10] It is to be noted, however, that the linear relation between d-c input voltage and frequency of the square-wave output voltage is quite unfavorable in those cases, e.g., design of test circuits for magnetic amplifiers, where it is necessary to vary the oscillation frequency of the magnetic-coupled multivibrator independently, without changing magnitude of the supply voltage.

In the adjustable-frequency magnetic-coupled multivibrator developed by R. L. Van Allen,[3] the oscillation frequency is controlled by means of a direct bias current applied to additional windings of a twin-core arrangement, similar to that of self-saturating magnetic amplifiers. In this case, the core flux instead of swinging between saturation levels,[2,4,5] is reset to an unsaturated value on alternate half-cycles. Van Allen's circuit, therefore, operates on variation of the core flux density in such a way that oscillation frequency can be controlled inversely proportional to the core flux. However, as a result of d-c bias magnetization of the cores, the output waveform deteriorates from its clean rectangular shape. This is undesirable in many applications of such d-c to a-c converters.

The author, when working with a differential-type magnetic-coupled multivibrator,[5] made the rather startling discovery that the oscillation frequency can be varied, within wide limits, without changing the magnitude of the output voltage and without impairing the output waveform, simply by connecting a saturating reactor with rectangular-hysteresis-loop core material in parallel to the load or across the collector terminals. Interconnection of a continuously-variable transformer made it possible to obtain linear frequency-control characteristics and to introduce coarse and fine adjustment with regard to both frequency and magnitude of the square-wave output voltage.

Since the anouncement of new circuits for magnetic-coupled multivibrators with wide-range frequency control,[11,12] further research and development work has been conducted toward the use of these circuits at high frequencies, up to about 100,000 cps. This paper reports on the results of these investigations.

Principle of Wide-Range Frequency-Control Circuits For Magnetic-Coupled Multivibrators

GENERAL

In the original form of the switching-transistor d-c to a-c converter introduced by G. H. Royer[2] and in the differential-type magnetic-coupled multivibrator described by C. H. R. Campling,[5] actual frequency of the square-wave output voltage is linearly proportional to the d-c input voltage. If a transformer with rectangular-hysteresis-loop core material is used and transistor switching occurs when the core saturates, the oscillation frequency is given by

$$f_0=\frac{E_{DC}}{4B_SA_TN_{eff}}\times 10^8 \qquad (1)$$

where E_{DC} is the d-c input voltage (volts), B_S is the saturation-flux density of the core material (gausses), A_T is the actual cross-sectional area (cm^2) of the transformer core, and N_{eff} is the effective number of turns of the transformer windings. In Royer's circuit, N_{eff} is the number of turns of the collector windings (N_2), whereas in Campling's arrangement N_{eff} is the difference (N_2-N_1) between the number of turns of collector windings and that of the emitter windings (N_1). On the other hand, in both circuits,[2,5] the output voltage E_{AC} is linearly proportional to the input voltage E_{DC}.

The basic principle of wide-range frequency-control circuits for magnetic-coupled multivibrators[11] consists essentially of varying, either step-by-step or continuously, the effective number of turns (N_{eff}) of windings which determine the frequency of oscillation in accordance with the relationship:

$$N_{eff}=\frac{E_{DC}}{4B_SA_Tf_0}\times 10^8 \qquad (2)$$

derived from equation 1. Because it is impractical to vary actual numbers of turns of the four winding units of the saturating transformer which operates in conjunction with the two switching transistors, the new method of wide-range frequency control employs additional components in the multivibrator circuit in

---

Paper 60-96, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., January 31–February 5, 1960. Manuscript submitted November 2, 1959; made available for printing November 25, 1959.

WILLIAM A. GEYGER is with the U. S. Naval Ordnance Laboratory, Silver Spring, Md.

The author wishes to thank E. T. Hooper, Jr. for valuable discussions of switching-transistor magnetic-coupled multivibrators and for his co-operation in conducting the oscillographic investigations.

Vol.
Pt. 1
196

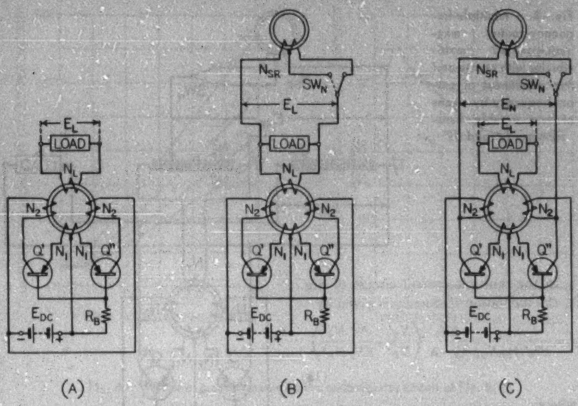


Fig. 1. The differential-type magnetic-coupled multivibrator circuit (A), and several forms of basic multivibrator circuits (B) and (C), providing wide-range frequency control

such a way that N_{eff} can be most readily varied through tapped windings and/or continuously-variable transformers.

Basic Circuits

In accordance with the schematic diagrams of Figs. 1 and 2, the evolution of new circuits may be characterized as follows:

In the differential-type arrangement, Fig. 1(A), as described by Campling,[5] the frequency of oscillation is given by

$$f_o=\frac{E_{DC}}{4B_S A_T(N_2-N_1)}\times 10^8 \qquad (3)$$

where $N_{eff}=(N_2-N_1)$ represents the effective number of turns of the magnetic-coupled multivibrator. The two p-n-p germanium transistors, Q' and Q'', are connected with collector windings N_2 and emitter windings N_1 of a saturating transformer with rectangular-hysteresis-loop core material and with a common-base resistor R_B limiting the base-to-emitter drive. The d-c input voltage E_{DC} may be supplied from a battery as illustrated in Fig. 1(A), or from a battery eliminator (power-supply transformer with full-wave rectifier and filter network).

When connecting a saturating reactor with rectangular-hysteresis-loop core material in parallel to the load terminals, as indicated in Fig. 1(B), the output frequency can be varied, within wide limits, without changing the magnitude of the output (load) voltage E_L and without impairing the output waveform by varying the effective number of turns N_{SR} of this saturating reactor (switch SW_N).[11] In this case, the peak voltage applied to the saturating-reactor winding N_{SR} will be

$$E_{SR}=E_{DC}\frac{N_L}{N_2-N_1} \qquad (4)$$

Provided that the product $B_S' A_{SR} N_{SR}$ of the saturating reactor is smaller than the product $B_S A_T(N_2-N_1)$ of the transformer, the output frequency will be solely determined by the parameters of this reactor and is then defined as

$$f_o=\frac{E_{SR}}{4B_S' A_{SR} N_{SR}}\times 10^8 \qquad (5)$$

where B_S' is the saturation flux density, A_{SR} is the actual cross-sectional area, and N_{SR} is the actual number of turns of the saturating reactor across N_L. In this special case, the output frequency is given by the equation

$$f_o=\frac{E_{DC}}{4B_S' A_{SR} N_{SR}(N_2-N_1)/N_L}\times 10^8 \qquad (6)$$

where $N_{eff}=N_{SR}\ (N_2-N_1)/N_L$ is the effective number of turns of the multivibrator circuit, as defined by equation 2. Evidently, in this special case, the frequency of oscillation is inversely proportional to N_{SR} and linearly proportional to N_L.

It is also possible to place the loading saturating reactor across an additional transformer winding or preferably across collector terminals, as shown in Fig. 1(C). In this case

$$N_{eff}=N_{SR}\frac{N_2-N_1}{2N_2} \qquad (7)$$

and the output frequency is given by

$$f_o=\frac{E_{DC}}{B_S' A_{SR} N_{SR}(N_2-N_1)/(2N_2)} \qquad (8)$$

Equation 6 makes it evident that linear frequency-control characteristics can be introduced by varying the actual number of turns of the transformer winding, across which the saturating reactor is connected. In the example of Fig. 2(A), this number of turns can be changed by means of a switch SW_E. When providing a load winding with numerous taps, the output frequency may be varied step-by-step, without changing

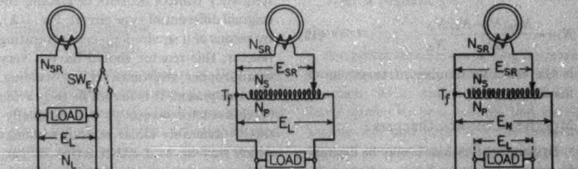


Fig. 2. Several forms of basic multivibrator circuits with linear frequency control, (A) with switch SW_E, (B) and (C) with continuously-variable transformer T_f

the magnitude of the output voltage and without impairing the output waveform. Of course, it is also possible to perform linear frequency control by varying actual number of turns of an additional secondary winding with properly distributed taps.

Fig. 2(B) illustrates a very convenient method for performing linear frequency control by means of a standard low-power adjustable transformer T_f connected across the load terminals. This simple arrangement makes it possible to vary the frequency continuously by moving the slider of T_f, i.e., by varying the turns ratio N_S/N_P, without changing the magnitude of output voltage E_L and without affecting the output waveform in an undesirable way. In this arrangement, the oscillation frequency is defined by the relationship

$$f_o=\frac{E_{DC}}{4B_S'A_{SR}N_{SR}}\frac{N_L}{N_2-N_1}\frac{N_S}{N_P} \tag{9}$$

where

$$N_{eff}=\frac{N_{SR}(N_2-N_1)}{N_L}\frac{N_P}{N_S} \tag{10}$$

is the continuously-variable effective number of turns of the multivibrator circuit; see equation 2.

In order to obtain greater flexibility, it is generally preferred[11] to connect the primary winding of the variable transformer T_f across the collector terminals, as indicated in Fig. 2(C). In this case, the output frequency is given by

$$f_o=\frac{E_{DC}}{4B_S'A_{SR}N_{SR}}\frac{2N_2}{N_2-N_1}\frac{N_S}{N_P} \tag{11}$$

where

$$N_{eff}=\frac{N_{SR}(N_2-N_1)}{2N_2}\frac{N_P}{N_S} \tag{12}$$

is the effective number of turns, as defined by equation 2.

MULTIPLE-CONTROL CIRCUITS

In some applications it may be desirable to provide coarse and fine adjustment of output frequency. Fig. 3 shows an example of a refined multivibrator circuit having coarse and fine adjusting transformers T_f' and T_f''. The large primary voltage for T_f' (coarse adjustment) is derived from the voltage E_N across both collector windings N_2. The small primary voltage for T_f'' (fine adjustment) is derived from a separate secondary winding N_T. The load is connected in this example across both collector windings. In this case, $E_N=2\ E_{DC}\ N_2/(N_2-N_1)$. Interconnected transformers may be used here in a similar way to provide coarse and fine adjustment of output voltage.[11]

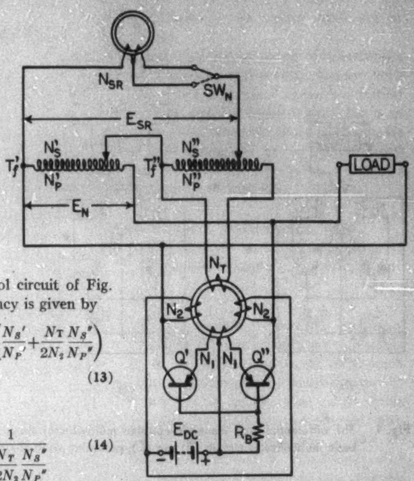


Fig. 3. Multiple-frequency-control magnetic-coupled multivibrator with coarse and fine adjustment of output frequency by means of two variable transformers, T_f' and T_f''

In the multiple-control circuit of Fig. 3, the oscillation frequency is given by

$$f_o=\frac{E_{DC}}{4B_S'A_{SR}N_{SR}}\frac{2N_2}{N_2-N_1}\left(\frac{N_S'}{N_P'}+\frac{N_T}{2N_2}\frac{N_S''}{N_P''}\right) \tag{13}$$

where

$$N_{eff}=N_{SR}\frac{N_2-N_1}{2N_2}\frac{1}{\dfrac{N_S}{N_P'}+\dfrac{N_T}{2N_2}\dfrac{N_S''}{N_P''}} \tag{14}$$

is the effective number of turns of the multivibrator circuit (equation 2). With the ratio $N_T/(2\ N_2)=1/10$, fine adjustment of output frequency can be achieved by varying the slider position of transformer T_f'', i.e., by varying N_S''.

Design Considerations

The more important aspects of proper design of the new multivibrator circuits of Fig. 1(B) and (C), Fig. 2(A) to (C), and Fig. 3 may be summarized as follows:

1. Because the fundamental principle of the described method for wide-range frequency control consists in loading the original differential-type circuit, Fig. 1(A), by means of a square-loop-core saturating reactor, this reactor should have a very small copper resistance of its winding. For this reason, it is favorable to provide on this reactor several groups of equally-rated (preferably bifilar-wound) windings which may be used either series- or parallel-connected to vary the oscillation frequency step-by-step over wide ranges.

2. In order to obtain optimized switching performance and to minimize "spikes" in the output square wave, magnetic-leakage effects must be reduced to the minimum. When using the turns ratio $N_1/N_2=1/2$, the transformer can be considered to have $N_1+2\ N_1+2\ N_1+N_1=6N_1$ turns acting as collector windings $(2N_1)$ and emitter windings (N_1) respectively. In this way of design, six twisted wires may be hand-wound simultaneously around the core of the transformer which then has negligible magnetic leakage between its winding units.

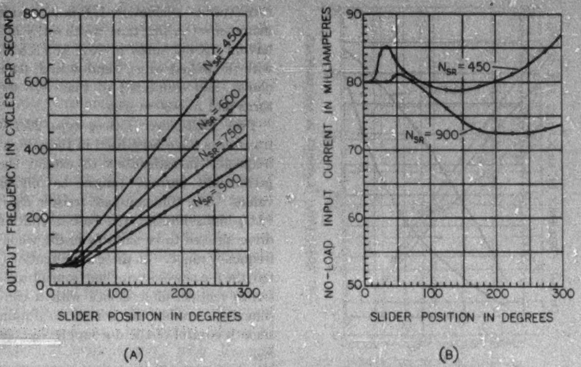
3. Although the transformer with windings N_1 and N_2 is always unsaturated with higher frequencies, it will operate with the lowest frequency between saturation levels, in accordance with the mode of operation of conventional multivibrator circuits, as shown in Fig. 1(A). It is, therefore, advisable to employ rectangular-hysteresis-loop core material also in this transformer, if the lowest part of the frequency range is to be utilized. However, when working only in the upper part of the frequency range, the transformer with N_1 and N_2 will never be saturated and will have correspondingly reduced core losses. In this case, lower-grade core material may be used in this transformer.

4. The continuously-adjustable transformer, Figs. 2(B) and (C), and Fig. 3, will, of course, never be saturated; and commercially-available components may be employed for this purpose. However, in order to minimize the magnetizing-current requirements of the adjustable transformer, it is necessary to use a standard low-power design. Of course, the scale of this transformer may be calibrated in cps for various values of d-c supply voltage E_{DC} and number of turns N_{SR} of the saturating reactor.

Performance Characteristics at Low Frequencies

DIMENSIONS OF COMPONENTS

A low-power (10-watt) prototype design with the circuit of Fig. 2(C) has been developed to be employed as a square



(A) (B)

Fig. 4. Results of measurements on the multivibrator circuit of Fig. 2(C)

A—Output frequency as a function of the variable-transformer slider position with the number of turns N_{SR} of the saturable reactor as a parameter

B—No-load input current as a function of the slider position with N_{SR} = 450 turns or 900 turns, respectively

Fig. 5 (below). Output waveform of the multivibrator circuit of Fig. 2(C) with:

A—Basic frequency 55 cps

B—Linear-range frequency 500 cps

C—Linear-range frequency 3,000 cps

wave generator for magnetic-amplifier transient analyzers[6–8] and magnetic-switch B-H loop tracers.[9,10] In these instrument applications, a very clean output waveform is of paramount importance to secure proper operation of the silicon-diode chopper circuits.

The transformer using an Orthonol 2-mil tape core (inside diameter = 1.5 in. (inches), outside diameter = 2.5 in., tape width = 1.0 in.) has 6×275 turns of No. *24* wire (6 twisted wires, hand-wound, 6×2.5 ohms) connected so that $N_1 = 275$ turns, and $N_2 = 550$ turns. The saturable reactor using Orthonol 2-mil tape core (inside diameter = $1^1/_4$ in., outside diameter = $1^5/_8$ in. tape width = 0.5 in.) has 6× 150 turns of No. *24* wire (6 twisted wires, hand-wound, 6×0.8 ohms), which may be used, series- or parallel-connected in various ways. Two germanium p-n-p transistors, type *2N158* (Q and Q''), are used in this design. The common-base resistor R_B limiting the base-to-emitter drive[5] is not critical ($R_B = 1{,}000$ ohms). The d-c power is supplied, in this case, from a transformer-rectifier-filter device with a continuously-variable transformer which permits smooth control of the d-c supply voltage E_{DC}.

FREQUENCY-CONTROL CHARACTERISTICS

In the prototype design of Fig. 2(C), both the output frequency f_o and the no-load input current I_{DC} have been measured as a function of the adjustable-transformer slider position, i.e., actual angular position (N_S) of the slider from its initial position ($N_S = 0$), with $N_{SR} = 450$, 600, 750, and 900 turns, respectively. Fig. 4(A) makes it evident that, within the initial range of from zero up to a certain slider position (e.g., 40 degrees with $N_{SR} = 900$), the saturable reactor (N_{SR}) remains unsaturated, and the transformer (N_1N_2) operates between saturation levels, in accordance with the conventional multivibrator circuit of Fig. 1(A). However, when moving the slider toward higher values of N_S, the saturable reactor starts to operate between its own saturation levels, while the transformer starts to have unsaturated mode of operation. That is, from zero up to a certain slider position (e.g., 40 degrees with $N_{SR} = 900$ turns), the transformer determines solely the output frequency (in this example, 55 cps). Then the saturable reactor takes over and represents the determining factor for the ranges 55 . . . 360 cps, 55 . . . 440 cps, 55 . . . 550 cps, and 55 . . . 740 cps, respectively. Thus, it is possible to perform smooth linear frequency control within these ranges.

Fig. 4(B) shows the no-load input current, I_{DC}, as a function of the variable-transformer slider position for $N_{SR} = 450$ turns, and $N_{SR} = 900$ turns. The considerable changes in magnitude of I_{DC} are caused by the fact that either the transformer (N_1N_2) is saturated, while the saturable reactor (N_{SR}) is unsaturated, or vice versa.

OUTPUT WAVEFORM

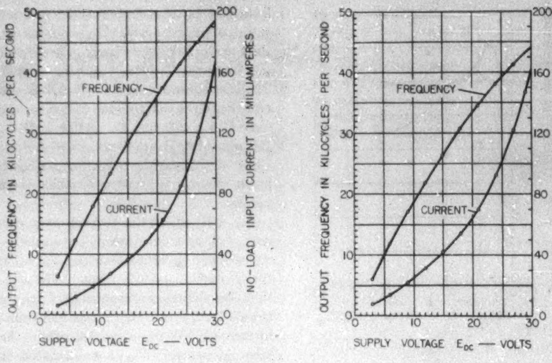
Fig. 5 presents oscillograms of the output voltage across the load resistor (output power, 10 watts) of the described prototype design for the "basic frequency" 55 cps (transformer N_1N_2 saturated, reactor N_{SR} unsaturated), and for the "linear-range frequencies" 500 cps and 3,000 cps (transformer N_1N_2 unsaturated, reactor N_{SR} saturated). With regard to the present applications in magnetic-amplifier test circuits[6–10] the spikes are sufficiently small. Moreover, if they are objectionable, they can be eliminated by connecting small capacitors (e.g., 0.05 microfarad) between collector and emitter of each transistor.[5]

Performance Characteristics at High Frequencies

GENERAL

A-c power supplies employing a square-loop-core saturating transformer in conjunction with switching transistors[2–5] have proved to be a suitable tool for reducing size and improving the performance of magnetic amplifiers by applying high power-supply frequencies in the range of 5 to 50 kc.[11–16] Recently, the author's investigations on high-frequency-operated low-level self-balancing magnetic amplifiers operating as audio amplifiers in conjunction with crystal-type radio receivers or standard turn-table equipment have shown that quite remarkable sound performance can be obtained with this arrangement. With regard to such special applications, the multivibrator circuits described in this paper have been investigated at high output frequencies, up to about 100 kc.

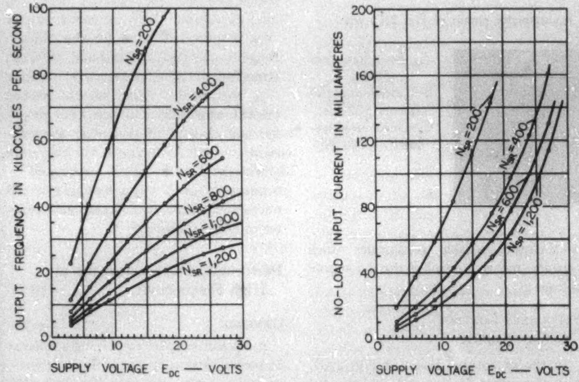
When applying magnetic-coupled multivibrators at high frequencies, eddy-cur-



(A) (B)

Fig. 6. Output frequency and no-load input current as a function of the supply voltage E_{DC}

A—For conventional circuit of Fig. 1(A)
B—For wide-range frequency control circuit of Fig. 1(C)



(A) (B)

Fig. 7. Results of measurements on the multivibrator circuit of Fig. 1(C) with $N_{SR}=200$ to 1,200 turns. As functions of the supply voltage E_{DC}

A—Output frequency
B—No-load input current

rent losses in the saturating core must be reduced to the minimum. This can be achieved by using magnetic bobbin cores made of very thin tapes.[17] In the present investigations concerning comparison of performance characteristics of the multivibrator circuits of Fig. 1(A) and Fig. 1(C) in the frequency range 5 to 100 kc, small 4-79 molybdenum Permalloy bobbin cores (tape width = 1/8 in., tape thickness = 1/2 mil, 1/4 mil, or 1/8 mil, respectively) have been used with the turns ratio $N_1/N_2=1/2$. Six twisted wires have been hand-wound on these cores with various numbers of turns to be employed in the circuit of Fig. 1(A) as collector windings ($N_2=2N_1$) and emitter windings (N_1). In the wide-range frequency control circuit of Fig. 1(C), the unsaturated transformer consisted of a small Orthonol 2-mil tape core (inside diameter = 1.0 in., outside diameter = $1\,^1/_4$ in., tape width = 1/8 in.) having 6×400 turns of No. *32* (B.&S.) wire (6 twisted wires, hand-wound, 6×8 ohms), and connected so that $N_1=400$ turns and $N_2=800$ turns.

Two high-power Delco, type *2N174* transistors have been used in these high-frequency investigations to extend the permissible current ranges to higher values. The common-base resistor $R_B=1{,}000$ ohms limiting the base-to-emitter drive[6] proved to be suitable in the whole frequency range. In these investigations, too, the d-c power is supplied from a transformer-rectifier-filter device with a continuously-variable transformer to obtain smooth control of the d-c supply voltage E_{DC}.

Results of Measurements

In order to compare the performance characteristics of the new circuit of Fig. 1(C) with those of the original differential-type circuit, Fig. 1(A), both the output frequency f_o and the no-load input current I_{DC} have been measured as a function of the d-c supply voltage E_{DC}. A calibrated vacuum-tube oscillator is used in the conventional way for measuring actual output frequency of the multivibrator circuit by means of Lissajous figures displayed on the screen of a cathode-ray oscilloscope.

Fig. 6(A) presents results of measurements on the original circuit, Fig. 1(A). The bobbin core (80 wraps of 1/8-mil tape) has 6×200 turns of No. *34* wire (6 twisted wires, hand-wound, 6×4 ohms, and connected so that $N_1=200$ turns and $N_2=400$ turns).

Fig. 6(B) shows results of correspondingly performed measurements on the new circuit of Fig. 1(C). The same bobbin core with $N_{SR}=4\times200=800$ turns has been used here as the loading saturating reactor, in conjunction with a small unsaturated transformer having an Orthonol 2-mil tape core (inside diameter = 1.0 in., outside diameter = $1\,^1/_4$ in., tape width = 1/8 in.) with 6×400 turns of No. *32* wire (6 twisted wires, hand-wound, 6×8 ohms, and connected so that $N_1=400$ turns and $N_2=800$ turns).

Comparison between Figs. 6(A) and 6(B) makes it evident that there is a relatively small increase in no-load input current I_{DC} as a result of the additional core losses in the unsaturated Orthonol-core transformer of the circuit of Fig. 1(C). When using this circuit, the output frequency f_o has slightly lower values than in the original circuit of Fig. 1(A).

Fig. 7 shows results of measurements on the circuit of Fig. 1(C) with $N_{SR}=200$

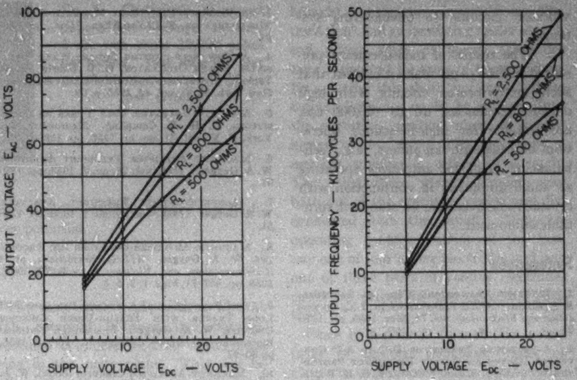
to 1,200 turns. Evidently, the output frequency f_o can be varied, within wide limits, by varying actual number of turns N_{SR} of the loading saturating reactor, the windings of the unsaturated transformer remaining unchanged. When decreasing N_{SR}, the no-load input current I_{DC} increases, correspondingly.

In order to determine power-handling capacity and efficiency of the circuit under test, Fig. 1(C), input and output values have been measured with load resistance, $R_L = 400$ to 2,500 ohms, connected across the collector terminals. Fig. 8(A) output voltage E_{AC} and Fig. 8(B) output frequency f_o as a function of the d-c supply voltage E_{DC} with load resistance R_L as a parameter. Fig. 9(A) makes it evident that, with $E_{DC} = 20$ volts, maximum output power, $E_{AC}^2/R_L = 6$ watts, is obtained in the circuit under test with $R_L = 600$ ohms.

The circuit efficiency is defined by the expression

$$\eta = \frac{\text{output power } E_{AC}^2/R_L}{\text{input power } E_{DC}I_{DC}} \times 100\% \qquad (15)$$

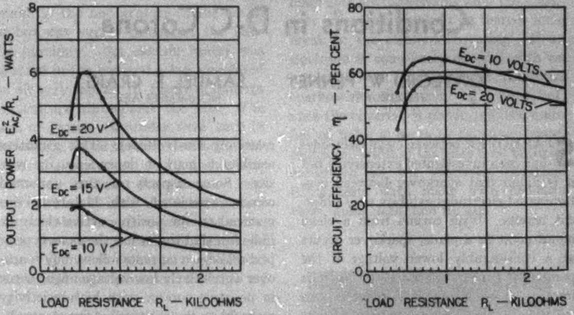
Fig. 9(B) shows $\eta = f(R_L)$ with E_{DC} as a parameter. With $R_L = 800$ to 1,000 ohms, the circuit efficiency is about 60% to 65%.



(A) (B)

Fig. 8. Results of measurements on the multivibrator circuit of Fig. 1(C) with $N_{SR} = 800$ turns. As functions of the supply voltage E_{DC}:

A—Output voltage
B—Output frequency



(A) (B)

Fig. 9. Results of measurements on the multivibrator circuit of Fig. 1(C) with $N_S = 800$ turns As functions of load resistance R_L with d-c supply voltage E_{DC} as a parameter:

A—Output power
B—Circuit efficiency

Output Waveform

The oscillograms of Fig. 10 show the output waveform of the circuit of Fig. 1(C) with $f_o = 10$, 50, and 100 kc. In the most interesting frequency range of 10 to 50 kc, a quite acceptable trapezoidal waveform is obtained, which is suitable for high-frequency operation of magnetic amplifiers.

Fig. 10. Output waveform of the multivibrator circuit of Fig. 1(C) with the frequencies 10 kc, 50 kc, and 100 kc

Telemetering Transducer Applications

Because output frequency of the new multivibrator circuits, Figs. 2(B) and (C), can be made directly proportional to angular position of an adjustable-transformer slider shaft, these circuits may be used in telemetering transducer applications. In this way, angular or linear displacement can be converted to a corresponding frequency shift. Movement or rotation of the shaft may be controlled remotely with a conventional servo system.

Conclusions

1. The output frequency of differential-type magnetic-coupled multivibrators can be varied, within wide limits, without changing the magnitude of the output voltage and without impairing the output waveform, simply by connecting a square-loop-core saturating reactor in parallel to the load or across the collector terminals.

2. Continuous linear frequency control can be achieved by introducing a standard low-power adjustable transformer connected across the load or across the collector terminals. This possibility suggests application of the new multi-

vibrator circuits to telemetering systems.

3. The results of measurements presented in this paper make it evident that such multivibrator circuits with high output frequencies, up to 100,000 cps, are suitable for high-frequency operation of magnetic amplifiers, e.g., self-balancing magnetic amplifiers operating as audio amplifiers in conjunction with crystal radio receivers or standard turntable equipment.

References

1. Junction Transistors Used as Switches, **R. L. Bright.** *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 74, Mar. 1955, pp. 111–21.

2. A Switching Transistor D-C to A-C Converter Having an Output Frequency Proportional to the D-C Input Voltage, **G. H. Royer.** *Ibid.*, July, pp. 322–26.

3. A Variable Frequency Magnetic-Coupled Multivibrator, **R. L. Van Allen.** *Ibid.*, pp. 356–61.

4. A New Self-Excited Square-Wave Transistor Power Oscillator, **G. C. Uchrin, W. O. Taylor.** *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 43, 1955, p. 99.

5. Magnetic Inverter Uses Tubes or Transistors, **C. H. R. Campling.** *Electronics*, New York, N. Y., vol. 31, no. 11, 1958, pp. 158–61.

6. Magnetic Amplifier Transient Analyzer, **W. A. Geyger.** *Ibid.*, vol. 26, no. 7, 1953, pp. 189–91.

7. Magnetic-Switch Transient Analyzer, **W. A. Geyger.** *Ibid.*, vol. 29, no. 1, 1956, pp. 150–51.

8. Magnetic-Amplifier-Operated Ink Recorders, **W. A. Geyger.** *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 77, Sept. 1958, pp. 457–71, Figs. 1, 2, 5, 6.

9. The Ferrotracer, A Simple Dynamic B-H Loop Tracer with Silicon-Diode Chopper Circuits, **W. A. Geyger.** *Proceedings*, National Electronics Conference, Chicago, Ill., vol. 12, 1956, pp.83–97.

10. Magnetic-Amplifier Circuits (book), **W. A. Geyger.** McGraw-Hill Book Company, Inc., New York, N. Y., second edition, 1957, pp. 350–67.

11. Frequency Control of Magnetic Multivibrators, **W. A. Geyger.** *Electronics*, vol. 32, no. 30, 1959, pp. 54–56.

12. Applications of Saturable Cores in Transistor Power Converters, **R. P. Putkovich, T. M. Corry.** *Proceedings*, National Electronics Conference, vol. 14, 1958, pp. 874–83.

13. An All-Magnetic Audio Amplifier System, **J. J. Suozzi, E. T. Hooper.** *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 74, July 1955, pp. 297–301.

14. High-Frequency Operation of Self-Saturating Magnetic Amplifiers, **H. W. Collins.** *Ibid.*, Sept., pp. 500–05.

15. Operational Magnetic Amplifier with Audio-Frequency Transistor Power Supply, **R. O. Decker, F. Gourash.** *Ibid.*, pp. 490–96.

16. Application of High Frequency Magnetic Amplifiers, **H. W. Collins.** *Proceedings*, National Electronics Conference, vol. 11, 1955, pp. 136–44.

17. Ultrathin Tapes of Magnetic Alloys with Rectangular Hysteresis Loops, **M. F. Littmann.** *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 71, July 1952, pp. 220–23.

Sparkover as Influenced by Surface Conditions in D-C Corona

GAYLORD W. PENNEY
FELLOW AIEE

SAMUEL E. CRAIG
NONMEMBER AIEE

SPARKOVER between clean electrodes has been investigated extensively but it is found that sparkover between contaminated electrodes exhibits very different results. With corona from a clean metal point to a plane, sparkover occurs at a considerably lower voltage if the point is positive.[1] Because of this, it is frequently assumed that positive corona will always result in the lowest sparkover voltage. A few experiments with contaminated electrodes show that sparkover is a complicated phenomenon and that negative corona does not always result in a higher sparkover voltage.

Some earlier work[2] indicates that, with positive corona, even small patches of insulating material on a wire could result in a condition called "abnormal corona," which frequently results in sparkover at relatively low voltages. This condition exhibits a marked dependence on wire size. Some aspects of this abnormal corona associated with high-resistivity material on the positive active electrode indicates that somewhat the same type of prebreakdown currents followed by sparkover at relatively low voltage might occur in negative corona with high-resistivity material on the positive passive electrode. Some preliminary experiments with high-resistivity materials on the passive electrode showed that this could result in marked reduction in sparkover voltage and that there seemed to be a correlation with resistivity.

The operating voltage of a precipitator is normally well below the sparkover voltage of a clean precipitator. This reduction in sparkover voltage is presumably due partly to dust suspended in the gas between electrodes and partly to dust on the electrodes. This paper is concerned only with the dust on electrode surfaces.

Flares

In this paper the term "flare" is applied to a type of discharge which appears to originate as a bright point on the anode and "flares" out in a cone-shaped, bluish glow, extending toward the cathode; see Fig. 1. Flares are particularly easy to produce with positive corona by putting high-resistivity material about 10 mils thick on the active electrode. The flare originates at an edge or hole in the insulating layer. While the point of the flare appears as bright glow, it does not ignite a material such as paper. It is only in case of sparkover that the paper burns. From this it appears that the glow is in the gas and that there is no high surface temperature.

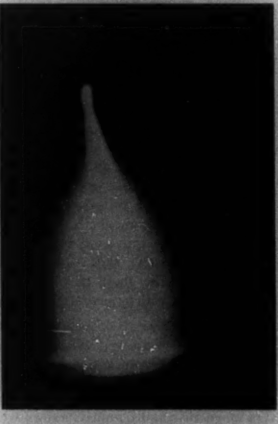


Fig. 1. Flare discharge: exposure 1/2 sec with 30,000 pulses per second; length of discharge 3 inches

---

Paper 60-103, recommended by the AIEE Electronics Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., January 31–February 5, 1960. Manuscript submitted September 30, 1959; made available for printing January 12, 1960.

Gaylord W. Penney and Samuel E. Craig are with the Carnegie Institute of Technology, Pittsburgh, Pa.

This work was supported in part by National Science Foundation grant *G2424* and was done in part by Samuel E. Craig in partial fulfillment of the requirements for the degree of Doctor of Philosophy at the Carnegie Institute of Technology.

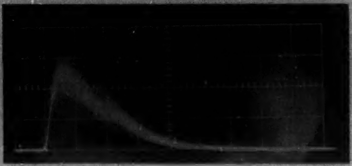


Fig. 2. Oscillogram of current pulse of flare discharge: exposure 3 sec with about 30,000 pulses per second. Time scale 0.1 μsec per division; current scale 18 ma per division

The current for a single flare has an average of 100 to 800 μamp (microamperes) but is made up of pulses having peak values of 10 to 60 ma (milliamperes). The pulse rises to a maximum in about 0.1 μsec (μsecond) and falls to less than 5% of the peak value in about 0.5 μsec. The time between pulses normally is about 33 μsec. The exact period between any two consecutive pulses may vary between 25 and 40 μsec. There is some random variation from one pulse to the next in both amplitude and time between pulses. The pulsation is not dependent on the external circuit or power supply. The pulse shape and its frequency of repetition are essentially the same regardless of the particular means by which the flare is produced; see Fig. 2.

These flares are easier to produce and are more stable in positive corona, but they can occur on the grounded electrode or anode in negative corona. Flares with negative corona are very important in precipitator operation since flares usually result in a very low flashover potential.

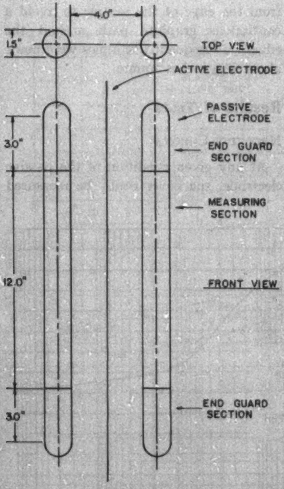


Fig. 3. Electrode dimensions used for sparkover tests

Apparatus

The tests were made with a wire at high potential between two parallel, grounded cylinders with all center lines lying in the same plane. This will be called the "wire between cylinders" arrangement. The arrangement was used because the discharge is relatively easy to see and photograph. The arrangement of electrode was enclosed in a box with a removable glass front.

Since the resistivity of most dusts and porous materials varies with the relative humidity, the temperature of the air in the box was varied to control the relative humidity. An electric heater was mounted in the box and a fan circulated the air very rapidly to insure a uniform temperature throughout the box. The large or passive electrodes were each 18 inches long. One of these electrodes consisted of a center section 12 inches long which was insulated from the end sections. This 12-inch insulated section was grounded through a meter for measuring the current per unit length of wire. The end sections served as guards and were grounded directly; see Fig. 3.

Early tests indicated that sparkover voltage depended on the resistivity of any coating on the anode. In the first apparatus with a single set of electrodes the resistivity was controlled by changing the relative humidity. But it takes a considerable time for it to reach equilibrium. Then, if the box is opened to make any change, it takes a considerable time for it to get back to equilibrium, and the relative humidity may not be exactly the same. To make it possible to change electrodes without disturbing the relative humidity, an apparatus was made with three pairs of passive electrodes and four active electrodes. These three electrode assemblies, shown in Fig. 4, were mounted in a box with the three shafts projecting through the box. Each shaft had a pointer outside the box. Thus, by rotating two shafts, any of the three pairs of cylindrical passive electrodes could be brought into the operating position. The third shaft had an insulating spider at each end. Four wires or active electrodes were supported by arms of these spiders. By rotating this shaft any one of these active electrodes could be brought into the operating position where it is connected to the high-voltage source by a member which also served as a latch. The other three wire electrodes were both out of the operating region and disconnected from the circuit. Thus, by operating controls outside of the box, any one of four active electrodes and any one of three pairs of passive electrodes could be placed in the operating position.

Since it is difficult to maintain a dust layer of uniform thickness, paper or cloth was used in most of the tests to cover the anode. Since such coverings tend to lower the sparking potential at some discontinuity in the covering, it is not necessary to cover the entire area. The electrode coverings on the electrodes in Fig. 4 show many dark spots where sparkover has occurred, but the number of these holes does not appreciably alter the sparkover potential provided that there is some break in the electrode covering. If an electrode is completely covered with an insulating layer having no holes, the initial sparkover voltage may be higher but, once the covering is punctured, the number of holes appears to be of no consequence. Most of the tests used paper or cardboard as the electrode covering. Some tests were made with ordinary cloth, glass cloth, and fly ash.

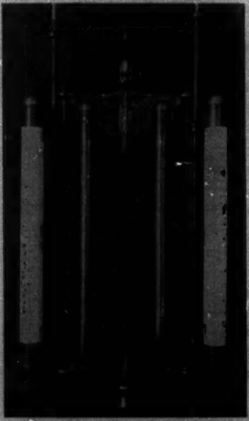


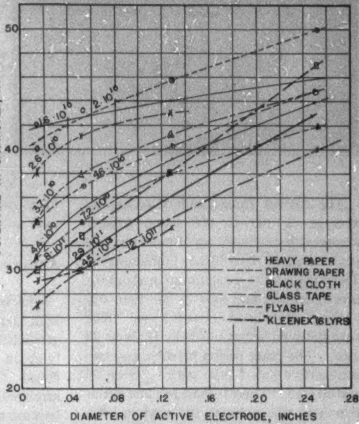
Fig. 4. Electrode assemblies

Measurement of Resistivity

The measurement of resistivity raises questions as to the equilibrium between the moisture content of the sample and the humidity of the air, particularly since the sample being tested is covered by the resistivity test electrodes. The test apparatus was designed to have the sample for the resistivity test in as nearly as possible the same condition as the material coating the electrodes. The resistivity tester was mounted in the same cabinet with electrodes. The upper electrode of the resisitivity tester could be moved by remote control so that one side of the sample was exposed to the atmosphere in the cabinet in the same manner as the electrode covering, except for the brief period when the resistivity was being measured and the upper test electrode was being pressed against the sample surface. The electrodes were each 2 inches square. The samples were very thin compared with 2 inches so that edge effects could be neglected.

First tests on the resistivity of heavy paper (cardboard 0.025 inch thick) showed that the current through the sample would commonly increase by a factor of 6 or 7 when the applied voltage was doubled. It was found that a much more linear relation between current and voltage was obtained when the sample was coated heavily with graphite where it contacted the electrodes. Apparently, for an uncoated sample, actual contact between the metal electrode and the sample occurs over only a fraction of the total surface, due to surface irregularities in the sample, and this fraction is increased by electrostatic force when the applied voltage is increased. This theory is substantiated by the fact that pressure on the electrodes affects the resistance through the uncoated sample, but pressure has no measurable effect when the sample is graphite coated. Subsequently, all paper samples were graphite coated where contact with a metal surface was involved. The graphite coating was kept about 1/4 inch away from the edge of the sample to avoid a continuous graphite path around the edge. Consequently samples were made about $2^1/_2$ inches square.

Fig. 5. Sparkover voltage as function of wire size for several passive electrode surface conditions with negative corona from a clean wire. Numbers at curves indicate resistivity of passive electrode coverings



Results of Tests

Negative Corona

At any given condition of the passive electrode, sparkover could be measured

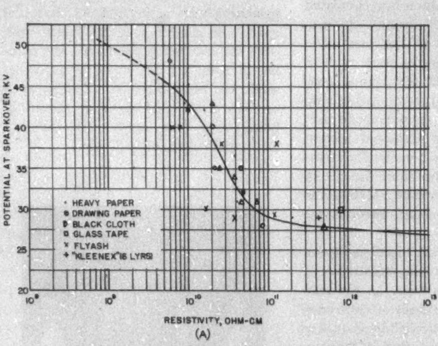
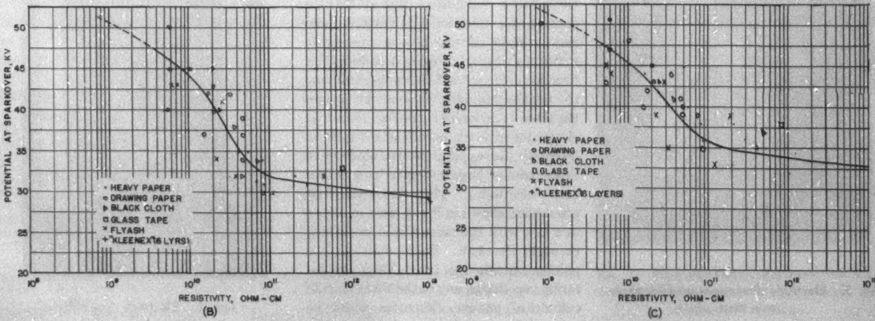


Fig. 6 Sparkover voltage as function of resistivity of passive electrode covering for negative corona from a clean wire. A—0.012-inch diameter. B—0.05-inch diameter. C—0.125-inch diameter



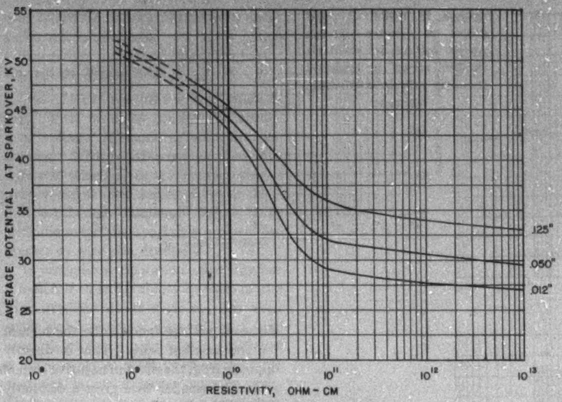


Fig. 7. Comparison of curves from Fig. 6

for four different sizes of active or high-voltage electrodes. Then, a pair of passive or large electrodes with a different covering could be brought into position and again sparkover measured with the same four active electrodes. The results of a number of such tests are plotted in Fig. 5. In this kind of test there is always considerable random variation. The curve for fly ash at 1.2×10^{11} ohm-cm (centimeters) lies below the curve for glass tape at 8×10^{11} ohm-cm and crosses the curves for heavy paper at 2.9×10^{11} ohm-cm and ordinary cloth at 4.5×10^{13}, but still fly ash at 2.6×10^{11} ohm-cm is up with the low-resistivity materials and the same fly ash at 1.2×10^{11} ohm-cm is down with the high-resistivity materials. All of the points for negative corona from a clean active electrode of 0.012-inch diameter are plotted as a function of resistivity in Fig. 6(A). While there is considerable scatter, there is a definite correlation with resistivity. Out of 31 points shown in Fig. 6(A) only four depart more than 10% of the flashover potential from the curve as drawn.

The results for a clean negative wire of 0.050-inch diameter are plotted in Fig. 6(B). In this case the correlation with resistivity is not as good, 11 out of 43 points lying at a potential more than 10% away from the curve as drawn. In spite of the scatter of the points the correlation with resistivity seems definite.

The curves of flashover voltage versus resistivity for three wire sizes are shown in Fig. 7. Because of the scatter and the number of conditions it was not considered feasible to show the points. In addition to the variation with resistivity, the flashover voltage decreases with wire size especially at higher resistivities.

In the case of negative corona with high-resistivity material on the active electrode, the results are essentially the same as when the active electrode is clean; see Fig. 8.

POSITIVE CORONA

In positive corona high-resistivity material on the passive electrode seems to have a somewhat smaller effect than in the case of negative corona. This is in agreement with the generally accepted idea that breakdown originates at the anode. Curves of breakdown potential versus wire size are shown in Fig. 9. In these cases breakdown is preceded by a high-current condition due to the back

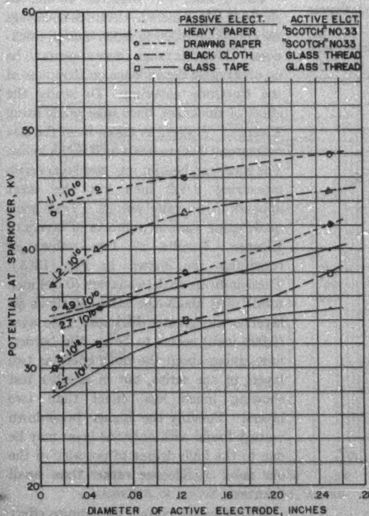
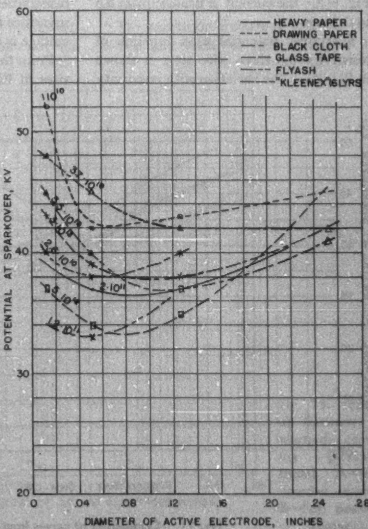


Fig. 8 (left). Effect of high-resistivity material on active electrode for negative corona. Numbers at curves indicate resistivity of passive electrode covering

Fig. 9 (right). Sparkover voltage as function of wire size for several passive electrode surface conditions with positive corona from a clean wire. Numbers at curves indicate resistivity of passive electrode covering



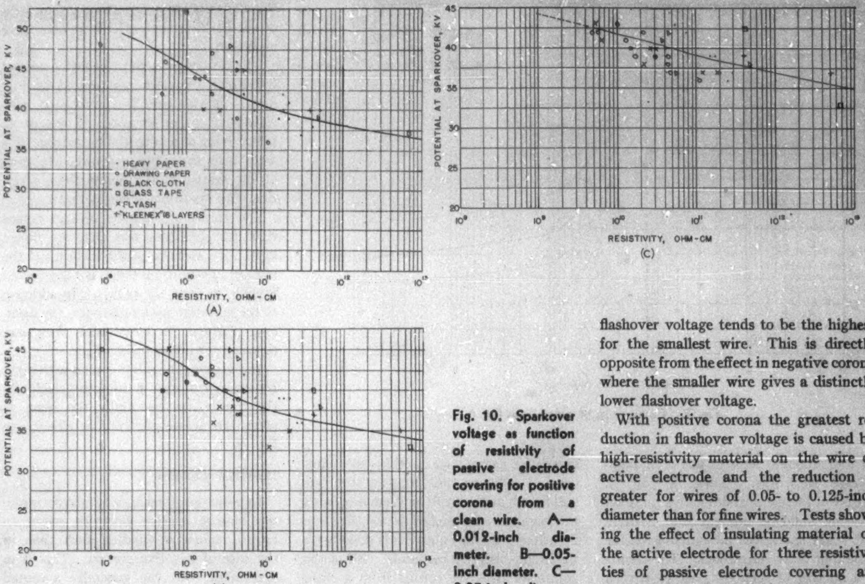


Fig. 10. Sparkover voltage as function of resistivity of passive electrode covering for positive corona from a clean wire. A—0.012-inch diameter. B—0.05-inch diameter. C—0.05-inch diameter

corona.[3] For the 0.012-inch wire the current may be as high as 5 ma per foot of wire prior to breakdown. For the comparable condition with negative corona Fig. 6(A), the current is usually from 0.5 to 0.8 ma.

Sparkover potential as a function of resistivity for three wire sizes is shown in Fig. 10. The variation with resistivity is not as great as in the case of negative corona but, still, the effects seems definite. With the 0.012-inch wire, Fig. 10(A) the highest flashover voltage is 37 kv for resistivities above 10^{12}, whereas for resistivities below 10^{10}, the lowest flashover voltage is 45 kv.

From the comparison of the three wire sizes in Fig. 11, it is evident that the flashover voltage tends to be the highest for the smallest wire. This is directly opposite from the effect in negative corona where the smaller wire gives a distinctly lower flashover voltage.

With positive corona the greatest reduction in flashover voltage is caused by high-resistivity material on the wire or active electrode and the reduction is greater for wires of 0.05- to 0.125-inch diameter than for fine wires. Tests showing the effect of insulating material on the active electrode for three resistivities of passive electrode covering are shown in Fig. 12. With insulating material on the wire or active electrode the flashover voltage is almost independent of resistivity of material on the passive electrode.

Effect of Thickness

The results for negative corona as shown in Fig. 6 include electrode coverings varying from 0.006 to 0.025 inch in thickness. Over this range there appears to be no observable dependence on thickness of the electrode covering. To study the effect of thickness, tests were made using Kleenex. A sheet as taken from the box measures 0.0032 to 0.004 inch in thickness. The sheet can be split into two layers each measuring 0.002 inch but the sheet appears to be quite porous. Tests were made using various numbers of layers of Kleenex. Results at two values of resistivity are shown in Fig. 13(A). These indicate that two layers (0.004 inch total) give about the same effects as six layers or any greater thickness. One layer (0.002 inch) seems to be on the borderline, giving about the same effect as two layers in one series, but in another test showing much less effect than two layers. Possibly the higher breakdown voltage found with only one layer may be due to the high degree of porosity of the one layer of Kleenex rather than small thickness.

With positive corona no definite effect

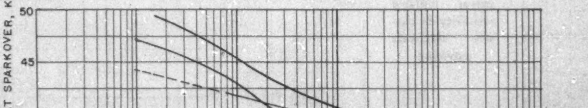


Fig. 11. Comparison of curves from Fig. 10

was discovered of thickness between 0.002 and 0.025 in thickness.

Effect of Wire Size

The effect of wire size is very interesting. With negative corona the sparkover voltage decreases as wire size decreases. With positive corona there is a wire size which gives a minimum sparkover voltage; this occurs between 0.05- and 0.125-inch diameter. Below 0.05-inch diameter the sparkover voltage increases as wire size decreases.

In precipitator operation a fine wire has a high self-cleaning ability because of the high gradient at its surface. But even with high-resistivity material on a fine positive wire, the reduction in sparkover voltage is not as great as when the same material is on a larger positive wire. High-resistivity material on a large positive wire usually produces flares, but the same material on fine wire will seldom result in a flare. In the present tests with fine wires, positive corona in general results in a higher sparkover potential than negative corona. However, for the wire sizes usually used in industrial precipitators, positive corona results in a distinctly lower sparkover potential than negative corona if there is any deposit on the wire.

Association of Flares with Sparkover

With high-resistivity material on a positive active electrode of 0.125-inch diameter, the flare type of discharge is particularly easy to produce and the flashover voltage especially low. The characteristic current pulses rise to a peak in 10^{-7} sec, and decay to a low value in approximately 5×10^{-7} sec. If the spacing to the passive electrode is four inches or greater it is relatively easy to produce a flare without accompanying sparkover. As the spacing is decreased, the probability of sparkover increases. At $1^1/_2$-inch spacing to the passive electrode, flares are almost always accompanied by frequent sparkover.

With negative corona the flares are associated with high-resistivity material on the passive electrode. Here they are much less stable and more likely to result in sparkover.

It appears that all cases of sparkover caused by high-resistivity material on anodes are preceded by these current pulses having about the same rise time. Of course, pulses having about the same rise and decay time have been observed by different experimenters and under somewhat different conditions. Parks[4] found higher current pulses having this characteristic rise and decay time. These were produced by applying a pulse between a sphere and plane. Other experimenters[5,6] have observed very similar pulses with clean electrodes at higher potential gradients. The characteristic rate of rise and decay of these current pulses would seem to indicate a common mechanism.

The repetition frequency and the rate of rise and decay of current appear to

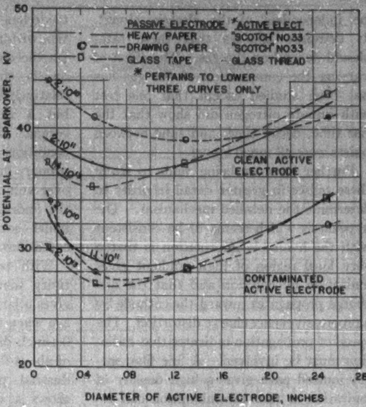
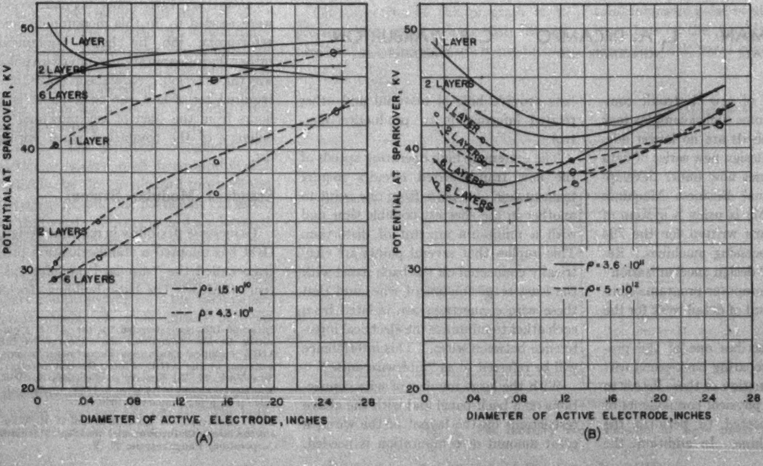


Fig. 12. Effect of high-resistivity material on active electrode for positive corona. Numbers at curves indicate resistivity of passive electrode covering



(A)

(B)

Fig. 13. Effect of thickness of electrode covering (Kleenex tissue). A—For negative corona. B—For positive corona

be independent of the electric circuit. It therefore appears that these rates should depend on the gas. The effect of varying the composition of the gas has not been studied except that one test with impure nitrogen did show that the rise time, decay time, and time between pulses were all much longer.

Further tests are being made to study these pulses in greater detail in order to determine the detailed mechanism. The hypothesis which seems to be consistent with the available information is that a very local region of high gradient gives ionization by collision and that the discharge propagates toward the cathode by a progressive space charge distortion. The speed of propagation indicates that radiation must be involved. A fine filamentary ionized path gives a high density of ionization so that recombination can account for the rapid decay of current. The present tests agree with Parks' tests in that if sparkover occurs there is a delay between the current pulse and the formation of a spark.

Conclusions

With negative corona there appears to be a definite correlation between sparkover voltage and the resistivity of any deposit or covering on the large or passive electrode. With positive corona, high-resistivity material on the passive electrode has a smaller effect and the correlation with resistivity is not as pronounced. On the contary, high-resistivity material on the small or active electrode has a relatively small effect with negative corona but a very large effect with positive corona. Thus it appears that the surface condition of the anode is the element of primary importance in determining sparkover voltage. This is in agreement with the findings of many other investigators under somewhat different experimental conditions.

With positive corona any high-resistivity material on the active electrode gives a large reduction in sparkover voltage when that electrode is of the order of 0.1-inch diameter. Thus with electrodes of about 0.1-inch diameter, as used in industrial precipitators, negative corona shows a considerably higher sparkover potential than positive corona. However, with a wire of less than 0.012-inch diameter as used in the low-voltage 2-stage precipitators, high-resistivity material on the positive active electrode has a much smaller effect, so that positive corona in general gives a higher sparkover potential.

While several types of high-resistivity materials were tested, it cannot be said that the tests are necessarily representative of all types of materials. It seems possible that some types exist which would not show the same correlation between resistivity and sparking. The effect of changes in atmosphere were not studied here; the tests reported were made with normal city air; the tests covered only a moderate range of absolute humidity.

These sparkovers at relatively low-voltage gradients are preceded by current pulses which characteristically rise to a peak in about 0.1 μsec and decay to a low value in 0.5 μsec. This rate of rise and decay of current seems to be a common characteristic of the prebreakdown currents observed by a number of investigators.

References

1. Der electrische Durchschlag von Luft zwischen konzentrischen Zylindern, **E. Uhlmann.** *Archiv für Elektrotechnik*, Berlin, Germany, vol. 23, 1929, p. 323.

2. Some Measurements of Abnormal Corona, **G. W. Penney, J. G. Hewitt, Jr.** *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 77, July 1958, pp. 319–27.

3. Electrostatic Precipitation of High-Resistivity Dust, **G. W. Penney.** *Ibid.*, vol. 70, pt. II, 1951, pp. 1192–96.

4. Surge Voltage Breakdown of Air in a Nonuniform Field, **J. H. Parks, H. N. Cones.** *Journal of Research*, National Bureau of Standards, Washington, D. C., vol. 56, no. 4, Apr. 1956, p. 201.

5. Electrical Breakdown of Gases, **J. M. Meek, J. D. Craggs.** Oxford University Press, London, England, 1953, p. 201.

6. Fast Time Analysis of Intermittent Point-to-Plane Corona in Air, **M. R. Amin.** *Journal of Applied Physics*, New York, N. Y., vol. 25 1954, pp. 210, 358.

Automation of Computer Panel Wiring

G. W. ALTMAN
NONMEMBER AIEE

L. A. DeCAMPO
NONMEMBER AIEE

C. R. WARBURTON
NONMEMBER AIEE

THE DESIGN of new electronic computers has become such a complex task that manual methods are no longer adequate. To help design new series of data processing machines now under development, International Business Machines Corporation (IBM) is using a system of computer programs written for the *704* and *705* data processing machines. Referred to as the "design mechanization" programs, these computer programs eliminate a large amount of detail work for the design engineer.

This paper describes one of the programs used for preparing back-panel wiring lists. The function of these lists is to describe how the physical components of circuits are connected to perform the logic of the machine. In addition, the lists provide bills of material and other related information to production control.

The extremely high operating speeds of the new transistorized circuits require transmission of a pulse from one point to another in the shortest possible time and with a minimum amount of distortion. This implies that several points are electrically connected on the back panel with the least total amount of wire, and that these wire connections are isolated from each other to minimize the electrical interference between wires. This interference will be referred to as "interwire noise."

With the large number of wire connections on a back panel and with the above restrictions on the layout of the wires, a great amount of computation is needed. Correlating these requirements now necessitates the use of a computer. Formerly, on an average, two engineering man-weeks were needed to do this monotonous and error-prone job by hand. Judgment rather than computation of interwire noise was the criterion. Now the average time required for a first run of a panel is 6 hours from the start of key punching to delivery of the completed panel wiring list.

Standard Modular System

To increase flexibility in machine design, IBM has adopted a standardization program called the "standard modular system" (SMS). The basic building blocks

---

Paper 60-128, recommended by the AIEE Computing Devices Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., January 31–February 5, 1960. Manuscript submitted August 3, 1959; made available for printing January 8, 1960.

G. W. Altman, L. A. DeCampo, and C. R. Warburton are with International Business Machines Corporation, Poughkeepsie, N. Y.

Vol
Pt.
196

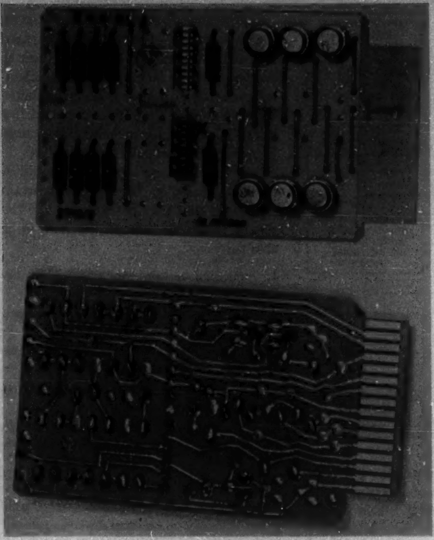


Fig. 1. Basic components of SMS printed circuit cards

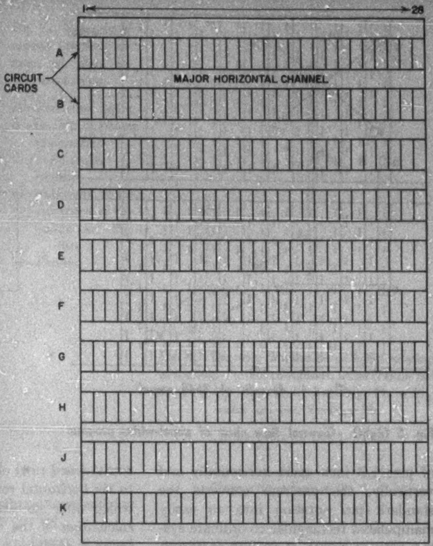


Fig. 2. SMS panel

in this system are printed circuit cards; see Fig. 1. The cards are plugged into a panel (Fig. 2) consisting of an array of sockets, numbered horizontally from left to right (1–28) and lettered vertically from top to bottom (A–K, excluding I). Pins protrude from the sockets to form a circuit card extension at the back of the panel; Fig. 3. The functioning of the circuits depends on how these pins are interconnected. The circuits that must be electrically common to perform a logical function possess certain pins which must be wired to each other. These pins as a group will be referred to as a "net." Each pin in the panel has a standard pin notation which represents its physical location. The notation designates the particular socket used horizontally and vertically by the circuit card to which the pin connects, and the particular pin letter for that card. Pins of a card are lettered from top to bottom (A–R, excluding I and O).

The matrix of pins, spaced 1/4 inch apart and projecting from the back panel, forms specific channels (spaces between rows) in which the wires must lie. Fig. 4 shows that these channels lie in four directions: horizontal (pin 6 to pin 7), vertical (pin 6 to pin 5), and two diagonals (pin 1 to pin 2 and pin 3 to pin 4). The horizontal channels fall into two categories, those that lie between: 1. horizontal rows of cards, referred to as major horizontal channels; and 2. horizontal rows of pins, referred to as minor horizontal channels. It can be seen from Fig. 4 that there is no straight-line route for a connection between pins 8 and 9.

General Description of Program

The objective of the program is to prepare back-panel wiring lists and a panel bill of material for use in production and assembly. Fig. 5 shows a general flow chart of the program. The wiring lists contain information about wire types, wire connections and their lengths, and wire routing. To accomplish this, the program accepts as input for multiple nets the pins that are electrically common, and determines the wire connections necessary to link these pins. It routes these connections in a simulated panel so that the total interwire noise of all the wires with the others on the panel is at a minimum. Since the back panel consists of a matrix

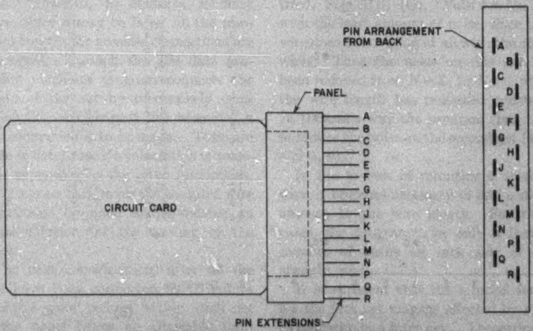


Fig. 3. Pin lettering and extension to back panel

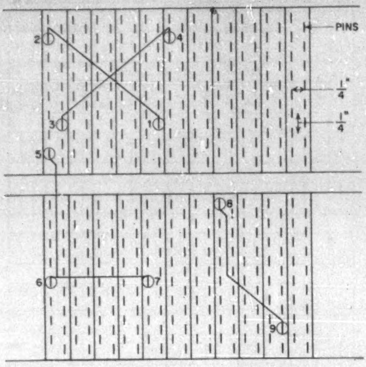

PINS
2
4
3
1
5
8
6
7
9
1/4"
1/4"

Fig. 4. Details of SMS panel

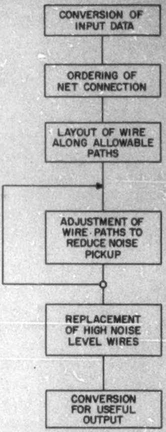

CONVERSION OF INPUT DATA
ORDERING OF NET CONNECTION
LAYOUT OF WIRE ALONG ALLOWABLE PATHS
ADJUSTMENT OF WIRE PATHS TO REDUCE NOISE PICKUP
REPLACEMENT OF HIGH NOISE LEVEL WIRES
CONVERSION FOR USEFUL OUTPUT

Fig. 5 (right). General flow chart of panel-wiring program

of pins 1/4 inch apart horizontally and vertically, the program converts the standard pin notation into an easily manipulated rectangular co-ordinate system with an eighth of an inch as its unit of length. Engineering tests have shown that the panel pins offer sufficient shielding for wires lying in adjacent channels. Hence, only interwire noise between wires which lie in the same channel and which overlap with each other needs to be considered. With the co-ordinate system, the program can consider the major horizontal (card row) channels as equivalent to six horizontal rows of imaginary pins which offer no shielding. This makes calculation of the noise interference between nonisolated wires in these channels practical.

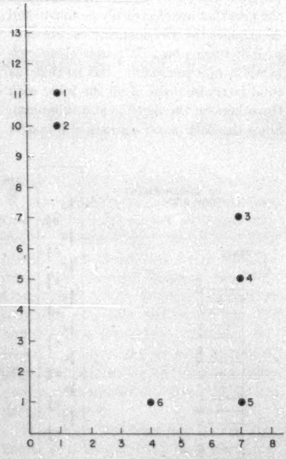
To do back-panel wiring, it is necessary that the *704* program receive certain information about the panel pins which are to be grouped together and made electrically common. This information, which is stored on magnetic tape and used as input to the program, is produced either from cards key-punched from lists prepared manually by the engineer, or from other computer programs of the design mechanization system. The latter programs use information received from the engineer to check the machine's logical design, to print lists of errors and component counts, and to formulate the standard notation for those pins that are to be connected on the back panel. The manually prepared lists of pins submitted by the engineer to do the same job lack the design mechanization system's ability to do error checking and to prepare bills of material, for instance, needed for machine production. However, the manually prepared lists do allow the engineer greater flexibility. He can specify to the program certain conditions that may be exceptions to the general rules; for example, his need to be able to:

1. preassign certain wire connections between pins, as well as the wire type used
2. designate pins that can only be used once
3. indicate certain wire types that must be used with particular pins
4. indicate certain wire types that must remain as he has preassigned them
5. indicate the maximum number of wraps (normally 2) that may be made to the pins of the panel

After the input has been translated into rectangular co-ordinate form, the program is ready to interconnect the pins of each net in an optimum manner, subject to production and engineering requirements.

Details of Procedure

Given *n* pins in a net, list in increasing order of length all the possible combinations of line segments which could link

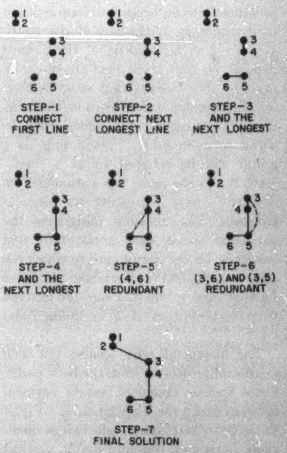


(A)

M	Line	R	M'
1	1, 2	1	1
2	3, 4	2	2
3	5, 6	3	3
4	4, 5	4	4
5	4, 6	5.9	16
6	3, 5	6	17
7	2, 3	6.7	7
8	1, 3	7.7	8
9	3, 6	7.9	9
10	2, 4	8.7	10
11	1, 4	9.7	11
12	2, 6	10.9	12
13	1, 6	11.9	13
14	2, 5	12.7	14
15	1, 5	13.7	15

R = Distance

(B)


1
2
3
4
6
5
STEP-1 CONNECT FIRST LINE
STEP-2 CONNECT NEXT LONGEST LINE
STEP-3 AND THE NEXT LONGEST
STEP-4 AND THE NEXT LONGEST
STEP-5 (4,6) REDUNDANT
STEP-6 (3,6) AND (3,5) REDUNDANT
STEP-7 FINAL SOLUTION

(C)

Fig. 6. Net used in example (A); combinations of n points taken two at a time, sorted into ascending order of distance (B); step-by-step solution of sample net (C)

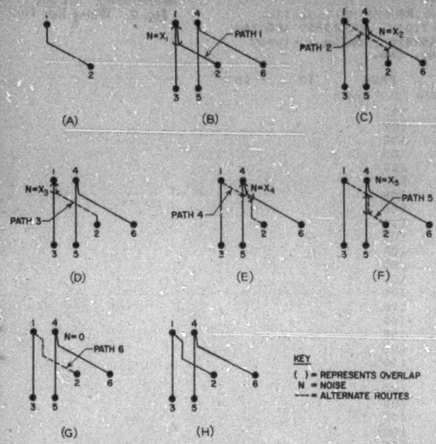


Fig. 7. Example of routing process; assume all wires are same type

A—Rerouting wire
B—Section of panel environment
C-F—Alternate routes
G—Alternate route accepted as solution at this time
H—Result

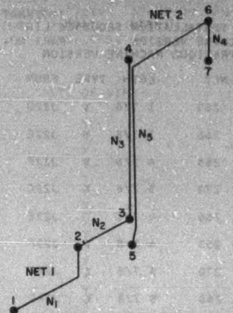


Fig. 8. Example of noise on an initial solution of routing. Initial conditions: noise (N) on net $1 = N_1 + N_2 + N_3 >$ limit where $N_3 > N_1$ or N_2; noise on net $2 = N_4 + N_5$

Connections for net 1: 1, 2
2, 3
3, 4
Connections for net 2: 5, 6
6, 7

this net. Then select the $n-1$ shortest line segments which do not contain any redundant lines or violate any rules.

The procedure can best be explained by an example. Suppose that the net shown in Fig. 6(A) is the one to be connected. All possible combinations of distances are computed and then sorted on increasing length, Fig. 6(B). In all cases the distances are computed only along the four allowable directions. A number M is then assigned to each of these distances. The steps to achieve a solution can be traced by referring to Figs. 6(B) and (C).

1. Bring out the first line (1, 2). Put its points in a "synonym table" (ST).
Assign $M^1 = M$. ST *1*——1,2

2. Bring out the next line (3, 4). Neither point is in ST *1*, so:
Assign $M^1 = M$. Generate ST *2*——3, 4

3. Bring out line (5, 6). Neither point is in ST *1* or ST *2*, so:
Assign $M^1 = M$. Generate ST *3*——5, 6

4. Bring out line (4, 5). Point *4* appears in ST *2* and point *5* in ST *3*. Therefore, ST *3* is added to ST *2*, and then ST *3* is erased. Hence:
Assign $M^1 = M$. ST *1*——1, 2
ST *2*——3, 4, 5, 6

5. Bring out the line (4, 6). Both points appear in the same synonym table, hence there is a redundant line. Assign M^1 equal to 1+(the maximum M). The synonym tables remain the same as in step 4.

6. Follow the same procedure as step 5 for line (3, 5).

7. Bring out line (2, 3). Point *2* appears in ST *1* and point *3* appears in ST *2*. Therefore ST *2* is added to ST *1* and then erased, leaving:
ST *1*————1, 2, 3, 4, 5, 6
ST *1* now has n(6) points, thus signifying that a solution has been reached.

8. Finally, select the $n-1$ lowest values of M^1, thus finding the $n-1$ lines which connect the n points.

1,2
3,4
5,6
4,5
2,3

Note that this method closely parallels the procedures of Loberman and Weinberger.[1] However, the procedures discussed in their paper were not completely applicable when the lines to a point were subjected to restrictions. In extending their procedures to include the previously mentioned restrictions, the authors of this paper do not guarantee an optimum solution.

For a given net, the method described would require n^2-n storage locations in excess of those required for the program itself. However, to conserve memory space, since n may be large, all the computed lengths for possible connections are not saved. Instead, the *704* data processing machines program connects the points of the net by successively computing the next shortest line whenever a new connection is to be made. Tests are made to determine if a calculation is necessary, depending on the given restrictions. The program then saves the accepted wire connections obtained and simulates an initial solution for the routing of the panel.

The next consideration, after all the nets have been connected, is to reduce interwire noise pickup within each net of the panel below an allowable limit determined by the engineer. At the same time, the program attempts to keep the length of each wire in the net the same as the length found initially. Using the initial solution for a particular wire as a starting point, alternate routes of equal length (in most cases there are several) are examined to find the one with the lowest noise level at this particular point of the computation. That route is then accepted as the new solution. The following relationship is used to compute interwire noise:

Noise = amount of wire overlap (inches) × noise-generating constant

Fig. 7 shows an example of the routing process. Assume that wire (1, 2) is the first one to be routed. All the equal length paths within the parallelogram are tried, Figs. 7(B)–(G). Path *6* is the one with the least amount of noise, since that wire does not overlap at all with the other wires. Thus the noise on this net has been reduced from $N = X_1$ to $N = 0$, while the wire length has remained constant. In the same way the program then tries to reduce the noise on the second net (4, 5, and 4, 6).

In the process of rerouting a straight wire, it becomes necessary to add a slight amount to the wire length. For these cases, the program tries only a limited number of paths on each side of the straight wire.

It is apparent that the solution found for the first net may be affected by a rerouted wire in a later net. Theoretically the program should resolve all the wires affected by a routing change. However,

```
                  INTERNATIONAL BUSINESS MACHINES
  INSTALLATION SEQUENCE LIST    PANEL 22B3    MACHINE  7101V   07-23-59  PG 26
MACHINE VERSION          PART NO.  0546835   ENG. CHG.  245782D   DATE     -  -
  PREVIOUS MACHINE VERSION        PART NO.              ENG. CHG.

  NET     LGTH  TYPE  FROM      VIA       VIA       TO        FROM      TO      P CH

   269    3 7/8   Y   J22D                          J23D

   266    3 7/8   Y   J22E                          J23E

   265    4 3/8   Y   J22F                          J24F

   270    3 7/8   Y   J23C                          J24C

   266    4 7/8   Y   J23E      J24G      J25G      J26E

   263    4 3/8   Y   J23F      J23E      J24E      J25F

   270    3 7/8   Y   J24C                          J25C

   268    3 7/8   Y   J24D                          J25D

   267    3 7/8   Y   J24E                          J25E

   265    4 3/8   Y   J24F                          J26F

   270    3 7/8   Y   J25C                          J26C

   268    3 7/8   Y   J25D                          J26D

   263    4 3/8   Y   J25F      J25E      J26E      J27F

   270    3 7/8   Y   J26C                          J27C

   268    3 7/8   Y   J26D                          J27D

   266    3 7/8   Y   J26E                          J27E

   183   12 6/8   T   A01B      A02D      B02D      C13D      A01J      C13J    T

   342   13 2/8   T   A01D      B01¤      C14D      C14F      A01C      C14J    T

   295   19 5/8   T   A01F      C26N      D26C      D27E      A01E      D27J    T

   325   14 6/8   T   A01H      A02F      A21F      A23A      A01G      A23J    T

   405   12 5/8   T   A01K      A01J      A18J      A19G      A01J      A19J    T

   311    5 2/8   T   A01M      A03G                A04G      A01L      A04J    T

   351    9 2/8   T   A01P                A07P      B12B      A01N      B12J    T

   441   13 7/8   T   A02H      A03K      A14K      B22D      A02G      B22J    T

   503   19 5/8   T   A11E      F11G      F10J      G10A

                            MAKE TEST WRAP
```

Fig. 9. Wiring list

because of machine time considerations, it was decided to allow the program to complete the routing of all the nets in the panel and then go back to the beginning and reroute the complete panel again. This was done to take into account the effect of all the rerouted wires on all the other wires in the panel.

The next step in the procedure is to test the computed noise level of each net in the panel against an allowable limit set by the engineer. If the noise level is excessive, the program reduces it below the limit by changing the type of wire used. Take as an example the configuration of Fig. 8, keeping in mind that there are many other wires in the panel which may have had an effect on this particular solution. Suppose that the total noise on all the wires in net *1* is greater than the allowable limit and that all the wires shown are single-strand. The program chooses the wire in net *1* with the highest noise level and changes it to a type of wire with the next lowest noise level as shown in the following breakdown. Thus the single-strand wire (3, 4) would be changed to a twisted pair. This procedure is repeated until the noise level specification is met for each net. Interwire noise decreases with each of the following changes in wire types:

1. Single-strand coupled to single-strand.
2. Single-strand coupled to twisted-pair.
3. Twisted-pair coupled to twisted-pair.
4. Single-strand coupled to coaxial (zero noise).
5. Twisted-pair coupled to coaxial (zero noise).
6. Coaxial coupled to coaxial (zero noise).

A secondary objective of the program is to alter previously computed panel wiring to accommodate engineering changes. To be considered a working tool by the engineer, it is necessary for the back-panel wiring program to incorporate any new additions and deletions to a panel without disturbing existing wires in their type or routing. The engineer can then investigate many variations on a particular panel, even override machine decisions under exceptional conditions, and receive results in time for evaluation. To production engineering and to the records-keeping area, the latest back-panel wiring list for a particular machine represents the most recent version of that panel.

The updating or change program for panel wiring closely parallels the methods used in the original wiring of a back panel. Input into the program may be from the design mechanization system as a result of a change or addition to logic design, or by a request from the

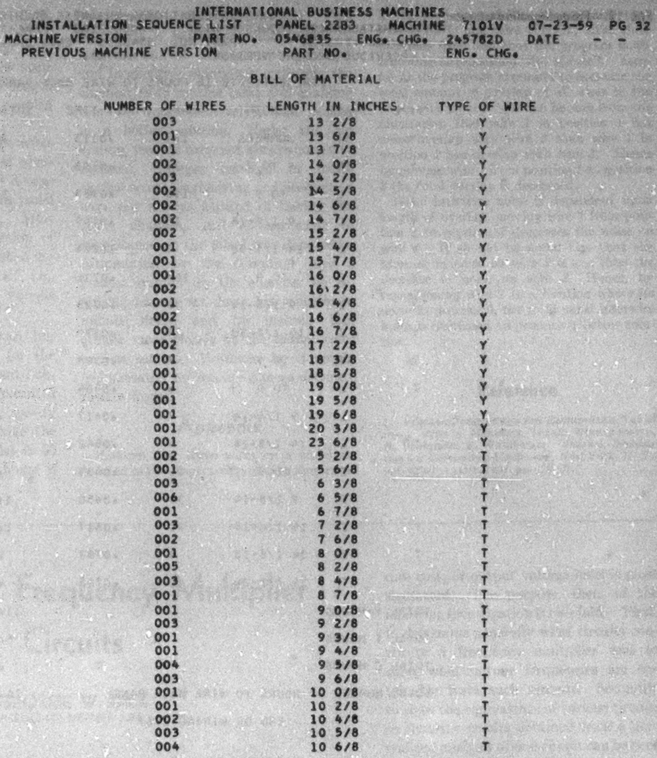
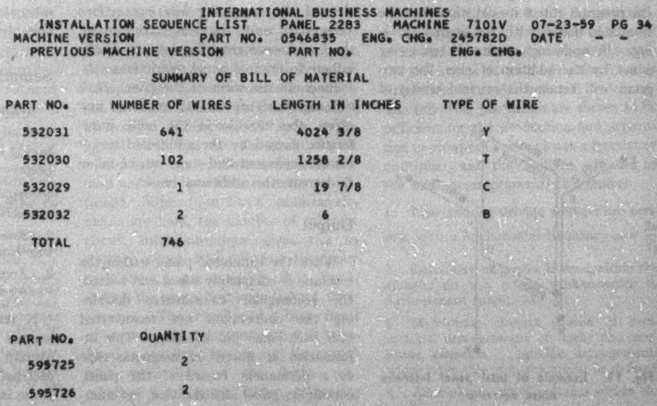


Fig. 10. Bill of material

```
                INTERNATIONAL BUSINESS MACHINES
   INSTALLATION SEQUENCE LIST    PANEL 2283    MACHINE  7101V   07-23-59  PG 32
MACHINE VERSION         PART NO. 0546835   ENG. CHG.  245782D   DATE      -  -
  PREVIOUS MACHINE VERSION        PART NO.               ENG. CHG.

                          BILL OF MATERIAL

           NUMBER OF WIRES     LENGTH IN INCHES     TYPE OF WIRE
                003                13 2/8                Y
                001                13 6/8                Y
                001                13 7/8                Y
                002                14 0/8                Y
                003                14 2/8                Y
                002                14 5/8                Y
                002                14 6/8                Y
                002                14 7/8                Y
                002                15 2/8                Y
                001                15 4/8                Y
                001                15 7/8                Y
                001                16 0/8                Y
                002                16 2/8                Y
                001                16 4/8                Y
                002                16 6/8                Y
                001                16 7/8                Y
                002                17 2/8                Y
                001                18 3/8                Y
                001                18 5/8                Y
                001                19 0/8                Y
                001                19 5/8                Y
                001                19 6/8                Y
                001                20 3/8                Y
                001                23 6/8                Y
                002                 5 2/8                T
                001                 5 6/8                T
                003                 6 3/8                T
                006                 6 5/8                T
                001                 6 7/8                T
                003                 7 2/8                T
                002                 7 6/8                T
                001                 8 0/8                T
                005                 8 2/8                T
                003                 8 4/8                T
                001                 8 7/8                T
                001                 9 0/8                T
                003                 9 2/8                T
                001                 9 3/8                T
                001                 9 4/8                T
                004                 9 5/8                T
                003                 9 6/8                T
                001                10 0/8                T
                001                10 2/8                T
                002                10 4/8                T
                003                10 5/8                T
                004                10 6/8                T
```



```
                INTERNATIONAL BUSINESS MACHINES
   INSTALLATION SEQUENCE LIST    PANEL 2283    MACHINE  7101V   07-23-59  PG 34
MACHINE VERSION         PART NO. 0546835   ENG. CHG.  245782D   DATE      -  -
  PREVIOUS MACHINE VERSION        PART NO.               ENG. CHG.

                SUMMARY OF BILL OF MATERIAL

PART NO.    NUMBER OF WIRES     LENGTH IN INCHES     TYPE OF WIRE

 532031           641               4024 3/8              Y

 532030           102               1258 2/8              T

 532029             1                 19 7/8              C

 532032             2                  6                  B

  TOTAL           746



PART NO.        QUANTITY

 595725             2

 595726             2
```

Fig. 11. Summary of bill of material

Fig. 12. Production time estimate

```
                          INTERNATIONAL BUSINESS MACHINES
      INSTALLATION SEQUENCE LIST    PANEL 22B3    MACHINE  7101V   07-23-59  PG 35
MACHINE VERSION          PART NO.  0546835   ENG. CHG.  245782D   DATE      -  -
  PREVIOUS MACHINE VERSION         PART NO.              ENG. CHG.

            TIME ESTIMATE IN HOURS TO WIRE BACK PANEL

TYPE   SHAPE    LENGTH      QUANT   UNIT TIME     TOTAL TIME

 Y      ST     JUMPERS       181      .0113          2.0453

 Y      ST      4 TO 9        39      .0163           .6357

 Y      L        TO 9        190      .0183          3.4770

 Y      L      9 1/8-14       18      .0199           .3582

 Y      L     14 1/8-30        8      .0324           .2592

 Y      Z        TO 9        147      .0216          3.1752

 Y      Z      9 1/8-14       35      .0223           .7805

 Y      Z     14 1/8-30       23      .0390           .8970

 T      ST       TO 9          1      .0398           .0398

 T      L        TO 9          6      .0404           .2424

 T      L      9 1/8-14        7      .0417           .2919

 T      L     14 1/8-24        4      .0542           .2168

 T      Z        TO 9         21      .0433           .9093

 T      Z      9 1/8-14       36      .0450          1.6200

 T      Z     14 1/8-24       21      .0617          1.2957

 T      Z     24 1/8-30        6      .0783           .4698

 C      ST    14 1/8-24        1      .0708           .0708

TOTAL Y HOURS                                       11.6281

TOTAL T HOURS                                        5.0857

TOTAL C HOURS                                         .0708

TOTAL NUMBER OF HOURS TO WIRE BACK PANEL            16.7846

                    END OF WIRING LIST
```

engineer to add or delete pins in an existing panel. Nets are processed one at a time to determine if they have been affected by the modification. If there have been no changes made to a net, the program enters its old wires into the simulated panel with its original routing. If updating results in enlarging a net by the addition of pins, the program will retain the original wiring of the net and its routing and connect the new pins to the old net by the same method described earlier. If updating results in reducing the number of pins in a net, the unwanted wires will be deleted, the remaining wire connections preserved, and the net reoptimized. Any new wires resulting from a change adhere to original panel restrictions.

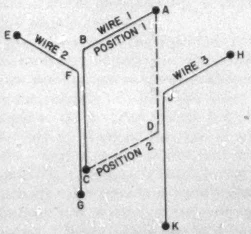


Fig. 13. Example of total panel interwire noise decrease

Since all the wires of the panel, both old and new, are entered into the machine, the increase in the noise interference caused by the additional wiring can be measured and steps can be taken to decrease this additional noise.

Output

When the simulated panel within the machine is completely wired and routed, the rectangular co-ordinates describing the connections are reconverted into back-panel pin notation. This information is stored on magnetic tape as a permanent record of the panel, containing panel identification, net number, wire length, wire type, and coded engineering data. Such invaluable aids in development and manufacturing as wiring lists, Fig. 9, bills of material, Figs. 10 and 11, and production time estimates, Fig. 12, can be extracted from this tape.

Summary and Conclusions

Briefly, the program initially minimizes over-all wire length on a panel while taking into account the following restrictions:

1. The maximum allowable number of connections to a pin.
2. Routing wires in channels to preserve wire length.
3. Accepting wire connections preassigned by the engineer.

It reroutes the wires, keeping their lengths constant to decrease the total overlap on the panel and hence to decrease total panel interwire noise. Noise is computed as follows:

$$\text{Noise (per net)} = L_1K_1 + L_2K_2 + \ldots + L_nK_n$$

where L_n = the length of overlap of a wire with another wire lying in the same channel and K_n = a function of length which depends on the wire types coupled together.

Essentially, by repositioning the wires into less dense channels, L (total overlap) is decreased, and by keeping K constant the total interwire noise in the panel is decreased; see the Appendix. After several iterations the final reduction of noise on each net is accomplished by keeping L constant and varying K, i.e., changing the wire types of the various wires that are coupled together.

The back-panel wiring program has produced to date 2,700 panels for the Poughkeepsie Product Development Laboratory groups. Previously, it required two engineering man-weeks to specify routings on a panel. By machine the average time to process a panel is 30 minutes of computer time and 6 hours of key punching and verification. This represents a saving in time (which is essential in all development projects) and in effective manpower (relieving the engineer from routine tasks and enabling him to work creatively), and results in a better solution. Under the old system, records were not always correctly updated, changes resulted in many clerical errors, engineering and production were not always assured of having the latest changes, and measurement of capacitance on the panel was impossible. Automation by the computer lays a greater burden on the engineer by requiring him to set down his problem in minute detail and by making him adhere more closely to the design rules he has set up. However, he is amply compensated by receiving more accurate results faster.

Appendix

Assume that three wires on a panel are located in the respective positions illustrated in the solid lines of Fig. 13. Assume that wire *2*(EFG) and wire *3*(HJK) have their positions fixed and that the program is attempting to reposition wire *1*(ABC). Insofar as the program attempts to decrease the total amount of overlap of all wires in the panel with all others, it can be seen from the illustration that wire *1* in position *1* has more overlap with wire *2* than wire *1* in position *2* has overlap with wire *3*. Hence by moving wire *1* from position *1* to position *2* the total overlap is decreased.

Since interwire noise is dependent upon length of overlap, moving wire *1* from position *1* to position *2* decreases the noise on wire *1*. It should be noted also that the increase in noise on wire *3* is less than the decrease in noise on wire *2*. Hence, by repositioning wire *1* to a location where its noise is decreased, the total panel interwire noise is decreased to produce a better solution.

Reference

1. FORMAL PROCEDURES FOR CONNECTING TERMINALS WITH A MINIMUM TOTAL WIRE LENGTH, H. Loberman, A. Weinberger. *Journal*, Association for Computing Machinery, New York, N. Y., vol. 4, no. 4, Oct. 1957, pp. 428–37.

Symmetrical Frequency Multiplier Circuits

GEORGE W. DICK
ASSOCIATE MEMBER AIEE

IN RECENT TIMES there has been considerable renewed interest in the use of static magnetic cores to produce power at a multiple of the frequency of a power source. Generally in these circuits, nonlinear magnetic reactors are connected in a symmetrical multiphase fashion, using the cores to produce harmonics and using symmetrical interphase connections to isolate desired groups of harmonics without the use of filters. In previous studies[1,2,3] several circuits of this type have been evaluated showing their characteristics when used either as a power supply or as a combined converter and power amplifier.

The synthesis and analysis of frequency multipliers of this type is complicated both by the presence of nonlinear elements and by the multiplicity of different linear network connections which can be used to isolate a particular group of harmonics. Examples of such topological variations are shown in Fig. 1 where only two nonlinear elements are used and even harmonic outputs are obtained. When from 2 to 6 nonlinear elements are used, the number of different circuit interconnections gives rise to roughly about 200 different frequency multipliers, many of which, as in Fig. 1, isolate the same harmonic output components. The selection of a best circuit of those in Fig. 1 for instance, will be related to economic factors and the particular application of the device, i.e., whether input power factor, saturating core cost, or output voltage level is most significant. The purpose, then, of the following investigation is two-fold. First, to determine generally what circuits constitute a frequency multiplier and to show what output frequencies are obtainable from such circuits. Secondly, to show the equivalence of various circuits so that the results obtained from a generalized analysis of one circuit can be used to evaluate several others.

General Discussion

A qualitative study of a general frequency multiplier circuit is proposed. By the application of methods of Fourier series and symmetrical component analysis, the synthesis and operation of this circuit and its equivalents will be defined. The following postulates are shown to be sufficient for the production and separation of groups of harmonics in a frequency multiplier, and they will be adopted as the basic assumptions of this study.

1. Time-varying driving voltages are periodic with a fundamental frequency $\omega_0 = \frac{2\pi}{T_0}$
2. Capacitive elements, if used within the system, do not create subharmonics or ferroresonant conditions.
3. Multiphase circuits consist of symmetrical arrangements of linear and nonlinear elements. Specific arrangements which have this property are to be defined.
4. All p-phase sets of voltages within the

Paper 60-66, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., January 31–February 5, 1960. Manuscript submitted June 22, 1959; made available for printing December 9, 1959.

GEORGE W. DICK, formerly with the University of Toronto, Toronto, Ont., Canada, is now with Bell Telephone Laboratories, Inc., Murray Hill, N. J.

This material represents a part of the author's doctoral thesis research at the Department of Electrical Engineering of the University of Toronto. The assistance of Prof. G. R. Slemon in supervising this project, is gratefully acknowledged. Both his and Prof. P. P. Biringer's suggestions have contributed to the theory and presentation of the paper.

The financial support of the Defense Research Board of Canada under Grant *4001–02* to the University of Toronto is also appreciated.

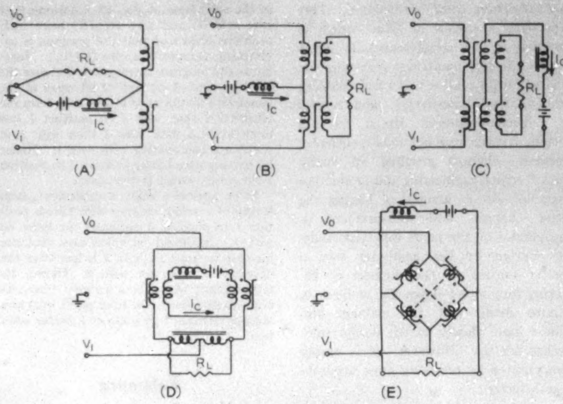


Fig. 1. **Equivalent frequency doubler circuits**

system, and including driving sources, are balanced in a time-sequential sense, i.e., if the voltages $v_0, v_1, v_2 \ldots v_{p-1}$ make up a balanced p-phase set

$$
\begin{aligned}
v_1(t) &= v_0(t - T_0/p)\\
v_2(t) &= v_0(t - 2T_0/p)\\
\vdots \quad & \quad \vdots\\
v_k(t) &= v_0(t - kT_0/p)\\
\vdots \quad & \quad \vdots\\
v_{p-1}(t) &= v_0(t - (p-1)T_0/p)
\end{aligned}
$$

A general circuit of this type is shown symbolically in Fig. 2. The circuit represents a tandem connection of symmetrical, multiphase series circuits, each composed of different sets of linear and nonlinear elements. Note that a nonspecific symbol has been chosen for the nonlinear element since it might represent a saturating core, diode, or other nonlinear electrical device. As a result of the basic assumptions it can be stated, regarding the nonlinear elements in any multiphase section of Fig. 2, that the currents through them as well as their terminal voltages must be time sequential in the steady state. Then without further reference to nonlinearities, except possibly to regard then as sources of harmonics, linear superposition methods may be applied to prescribe symmetrical phase-reducing networks between sections of specific circuits. Accompanying each phase reduction there is shown to be a related isolation of harmonic components.

Before proceeding it is worthwhile to note the significance of assumption 4 regarding multiphase voltages. When only fundamental and odd-harmonic voltage components are present, the conventional single-phase 3-wire, or Edison system, represents what is defined as a time-sequential system 2-phase system. Generally, however, there is a definite distinction between time-varying single-and 2-phase balanced systems. Fig. 3 shows that it is not possible to consider the single-phase voltage $V_{01}(t)$ as a time-sequential 2-phase system of voltages $V_{0n}(t)$ and $V_{1n}(t)$. Similar considerations must be applied if it is desired to convert a conventionally termed 2-phase source of harmonic voltages to a balanced 4-phase system.

Phase Sequential Harmonic Components

As a result of the previous basic assumptions the harmonic components of a p-phase set of voltages or currents can be defined in terms of the corresponding components at a reference node or branch of the set. Considering $v_0(t)$ a reference node voltage.

$$v_0(t) = \sum_{\omega=-\infty}^{\infty} V_0(j\omega) e^{j\omega t} \tag{1}$$

where

$$V_0(j\omega) = \frac{1}{T_0} \int_0^{T_0} v_0(t) \epsilon^{-j\omega t}\, dt \tag{2}$$

Since $V_0(-j\omega) = \bar{V}_0(j\omega)$, a complex or vectorial representation of $v_0(t)$, can be made by considering only the positive frequency components in the complete Fourier series of equation 1.

Let

$$V_0(t) = \sum_{\omega=0}^{\infty} V_0(j\omega) \epsilon^{j\omega t} \tag{3}$$

where

$$V_0(j0) = \frac{1}{T_0} \int_0^{T_0} v_0(t)\, dt \tag{4}$$

$$V_0(j\omega) = \frac{2}{T_0} \int_0^{T_0} v_0(t) \epsilon^{-j\omega t}\, dt \quad \text{For } \omega \neq 0 \tag{5}$$

hence

$$v_0(t) = \frac{1}{2} (V_0(t) + \bar{V}_0(t))$$

Then utilizing assumption 4 the harmonic components of $V_1(t)$ can be related to those of $V_0(t)$

$$
\begin{aligned}
V_1(j\omega) &= \frac{2}{T_0} \int_0^{T_0} v_0(t - T_0/p) \epsilon^{-j\omega t}\, dt\\
&= \epsilon^{-j\omega T_0/p}\, V_0(j\omega)
\end{aligned}
$$

Let

$\omega = q\omega_0$, where $q = 0, 1 \ldots \infty$

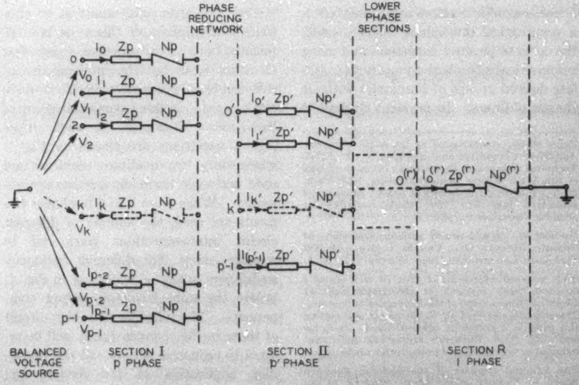


Fig. 2. **General frequency multiplier tandem configuration**

Vol
Pt.
196

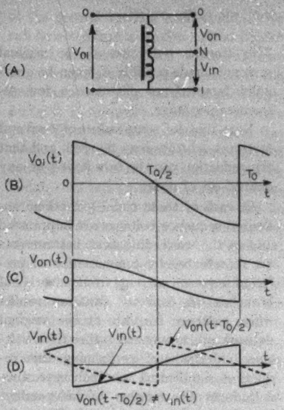


Fig. 3. Distinction between one- and two-phase time sequential voltages

$$V_1(j\omega)=\epsilon^{-jq\,2\pi/p}\,V_0(jq\omega_0)$$

Similarly $V_k(j\omega)$ referring to phase k

$$=\epsilon^{-jkq\,2\pi/p}\,V_0(jq\omega_0)$$

Since

$$\epsilon^{-jkq\,2\pi/p}=\epsilon^{-jk(q+np)2\pi/p}$$

for integers q, and n, by allowing q to take the values 0, 1, 2 … p-1, and n to take all positive integral values 0, n each harmonic component is defined by a different (q, n) and the complex Fourier series for $V_k(t)$ can be sectioned into p infinite subseries each of which corresponds to a particular q, and is related to the corresponding subseries of $V_0(t)$ by an individual phase operator

$$\epsilon^{\,jkq\,2\pi/p}$$

i.e.

$$V_k(t)=1\sum_{n=0}^{\infty}V_0(j(0+np)\omega_0)\epsilon^{j(0+np)\omega_0 t}+$$
$$\epsilon^{-jk1\cdot 2\pi/p}\sum_{n=0}^{\infty}V_0(j(1+np)\omega_0)$$
$$\epsilon^{j(1+np)\omega_0 t}\ldots+\epsilon^{-jkq\,2\pi/p}\sum_{n=0}^{\infty}V_0(j(q+$$
$$np)\omega_0)\epsilon^{j(q+np)\omega_0 t}\ldots+\epsilon^{-jk(p-1)2\pi/p}$$
$$\sum_{n=0}^{\infty}V_0 j(p-1+np)\epsilon^{(p-1+np)\omega_0 t} \qquad (6)$$

Each of these subseries of $V_0(t)$ represent a complex time functional component denoted by $V_{0q}(t)$ which may be referred to as a qth sequence component of $V_0(t)$

$$V_{0q}(t)=\sum_{n=0}^{\infty}V_0(j(q+np)\omega_0)\epsilon^{-j(q+np)\omega_0 t} \qquad (7)$$

and the real-time functional component is given by

$$v_{0q}(t)=\frac{1}{2}[V_{0q}(t)+\bar{V}_{0q}(t)]$$

where

$$\bar{V}_{0q}(t)=\sum_{n=0}^{\infty}V_0(-j(q+np)\omega_0)\epsilon^{+j(q+np)\omega_0 t}$$

Then

$$V_0(t)=\sum_{q=0}^{p-1}V_{0q}(t) \qquad (8)$$

$$v_0(t)=\sum_{q=0}^{p-1}v_{0q}(t) \qquad (9)$$

It is important to note that the sequence components $v_{0q}(t)$ are orthogonal time functions because any two of these obtained for different values of q do not contain similar elementary harmonic functions.

Since

$$\int_0^{T_0}V_0(j\omega_1)\epsilon^{j\omega_1 t}\,V_0(j\omega_2)\epsilon^{j\omega_2 t}dt=0 \text{ for } \omega_1\neq\omega_2$$

Then

$$\int_0^{T_0}v_{0q_1}(t)v_{0q_2}(t)dt=0 \text{ for } q_1\neq q_2$$

This property is used in a later section concerned with the design of circuit components on the basis of rms currents or voltages. It implies

$$[V_0(\text{rms})]^2=\sum_{q=0}^{p-1}[V_{0q}(\text{rms})]^2 \qquad (10)$$

From equation 6 the sequence components of $V_k(t)$ can be expressed in terms of those of $V_0(t)$

i.e.

$$V_{kq}(t)=\epsilon^{-jkq\,2\pi/p}\,V_{0q}(t) \qquad (11)$$

Giving

$$V_k(t)=\sum_{q=0}^{p-1}\epsilon^{-jkq\,2\pi/p}\,V_{0q}(t) \qquad (12)$$

The relationship between node voltages and the sequence components of the reference phase is conveniently shown in matrix notation by

$$[V_k(t)]=[\epsilon^{-jk,q\,2\pi/p}][V_{0q}(t)] \qquad (13)$$

The inverse relationship also must exist and is

$$[V_{0q}(t)]=[\epsilon^{+jk,q\,2\pi/p}][V_k(t)] \qquad (14)$$

Similar relationships will apply to balanced p-phase branch currents.

Synthesis of a Tandem Circuit

Apparently expressions 13 and 14 are identical to those used in fundamental frequency, symmetrical component analysis. It is noted however, that each of the sequence components described here are summations of vectorical harmonic components. If these vectors were known at any given time the transformation would give the value of the phase voltages. The expressions are used in this study since they give a single-phase frequency-domain interpretation of the time-sequence condition for multiphase voltages or currents.

First equation 13 can be used to define a symmetrical phase reducing network for the p-phase section of the general frequency multiplier circuit. In the following phase transformation for $p=6$ columns of the matrix indicate that certain sequence components occur at different nodes k in identical phase, i.e., for $q=0$, 2, 3, and 4.

$$\begin{bmatrix}V_0(t)\\V_1(t)\\V_2(t)\\V_3(t)\\V_4(t)\\V_5(t)\end{bmatrix}=\begin{bmatrix}1&1&1&1&1&1\\1&b&b^2&b^3&b^4&b^5\\1&b^2&b^4&1&b^2&b^4\\1&b^3&1&b^3&1&b^3\\1&b^4&b^2&1&b^4&b^2\\1&b^5&b^4&b^3&b^2&b\end{bmatrix}\begin{bmatrix}V_{00}(t)\\V_{01}(t)\\V_{02}(t)\\V_{03}(t)\\V_{04}(t)\\V_{05}(t)\end{bmatrix}$$

where

$$b=\epsilon^{-j2\pi/6}$$

The zero sequence components will always be in phase for all lines regardless of the value of p. A symmetrical p to 1 phase reduction can always be made by joining all k nodes to a common $0'$ node. The effect of such a nodal reduction is to reduce the number of independent phase voltages to 1 and from equation 14 it can be shown that all sequence voltage components at the original k nodes are eliminated except for those of the zero sequence. The time sequence condition is still applicable to these nodes and the system remains symmetrical.

In general other nodal reductions will be possible provided p can be factored by any integer p'. In this case some qth sequence voltage component q_1 where $q_1 p'=p$, will be present at nodes

$$k,\ k+p',\ k+2p'\ldots k+(q_1-1)p'$$

in equal phase since

$$\epsilon^{-jkq_1\,2\pi/p}=\epsilon^{-j(k+p')q_1\cdot 2\pi/p}\ldots\text{etc.}$$

Note that the zero sequence is also present at these nodes with equal phase, and so is the q_1th sequence component if

$$q_1 p'=2p,\ 3p\ldots\text{or }(p'-1)p$$

In order to obtain a symmetrical reduc

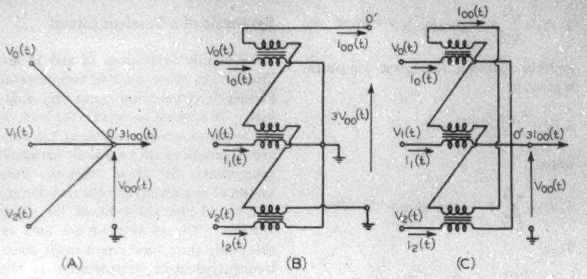


Fig. 4. Linear nodal and branch reducing networks for $p=3$, $p'=1$

A—Nodal reduction
$V_0(t)=V_1(t)=V_2(t)=V_{00}(t)$

B—Branch reduction
$I_0(t)=I_1(t)=I_2(t)=I_{00}(t)$

C—Alternative branch reduction
$I_0(t)=I_1(t)=I_2(t)=I_{00}(t)$

tion with the suppression of all other sequence voltages, all nodes having the sequence components q_1 in equal phase must be treated similarly. Where

$$q_1p'=0,\ p,\ 2p\ldots(p'-1)p \qquad (15)$$

Thus

node 0 is joined to node $0'$
node 1 is joined to node $1'$
$\vdots$
node $p'-1$ is joined to node $(p'-1)'$
node p' is joined to node $0'$
node $p'+1$ is joined to node $1'$
$\vdots$
node $p-1$ is joined to node $(p'-1)'$

This has produced a balanced p-phase set of node voltages containing sequence components given by equation 15. The reduced phase system is time sequential on both a p-phase basis and a p'-phase basis if it is understood that each new node or branch represents more than one of those of the original system. A p'-phase set of equal linear and nonlinear components can be included in this section followed by another nodal phase reduction to some factor p'' of p'. Sequence voltage components of the following p'' phase system are given on the original p-phase basis by

$$q_1p''=0,\ p,\ 2p,\ \ldots(p''-1)p$$

This treelike reduction of the original system is made possible if p can be gradually factored to eliminate a few sequences per stage.

Note that for some values of p several factoring schemes may be used, and that the reduction can be terminated at any desired set of nodes.

At each of these nodel reductions undesired sequence voltages are suppressed and in the succeeding branches currents are represented by a summation of preferred sequence currents from the original branches. A converse situation results when sequence currents of the original network are suppressed rather than voltages. Methods of performing this reduction, signified as a branch type since it reduces the number of independent branch currents, are shown in Fig. 4 for $p=3$, $p'=1$.

A modified reduction network such as shown in Fig. 4(C) using an ideal, nonsaturating, core-type trans-former is necessary when zero-sequence direct currents are to be propagated. In the general frequency multiplier circuit either nodal or branch reduction networks can be chosen to connect any two multiphase sections. Branch reductions allow elimination of sequence current components from the driving source and first section

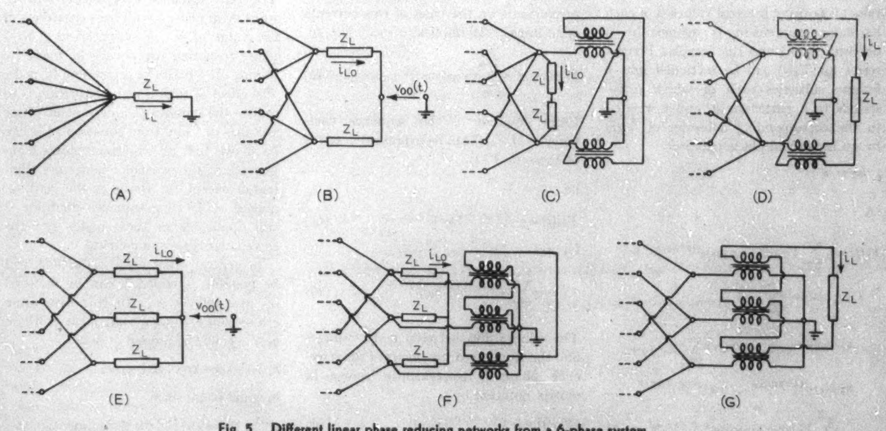


Fig. 5. Different linear phase reducing networks from a 6-phase system

A—$p'=1$
$i_L=6i_{00}(t)$

B—$p'=2$, $p''-1$
$i_{Lo}=3i_{03}(t)$
$i_{00}(t)=0$

C—$p'=2$, $p''=1$
$i_{Lo}=3i_{03}(t)$
$V_{00}(t)=0$

D—$P'=2$, $p''=1$
$i_L=3i_{00}(t)$
$i_{03}=0$

E—$p'=3$, $p''=1$
$i_{Lo}=2(i_{02}(t)+i_{04}(t))$
$i_{00}(t)=0$

F—$p'=3$, $p''=1$
$i_{Lo}=2(i_{02}(t)+i_{0}(t))$
$V_{00}(t)=0$

G—$p'=3$, $p''=1$
$i_{0L}=2i_{00}(t)$
$i_2(t)=i_{04}(t)=0$

of the general tandem configuration. It can be noted from equation 15 that fundamental frequency (first sequence) currents would always be eliminated if the first phase reduction were of branch type. Since it is normally desired to convert power from the fundamental, source frequency, to load power in higher sequence components, the first reduction must be nodal. All other possible sequence isolations for a given p-phase system can be determined empirically by performing alternative nodal or branch reduction schemes on the tandem configuration.

It is noted that the use of nonlinear elements after the first section of the general circuit of Fig. 2 can effect quantitative changes in the various sequence components but cannot isolate these in a manner different from a purely linear-phase-reducing network. In the following discussion only one p-phase of set nonlinear elements will be referred to, with succeeding p', p'', ... phase sections composed of linear elements. Qualitative results then obtained could be extended to fit the more general case. Examples of linear phase reductions are given for $p=6$ in Fig. 5.

EQUIVALENT CIRCUITS

It was indicated at the beginning that many magnetic-core frequency multiplier circuits could be used to obtain a given frequency output. Such circuits, if symmetrical, will be mathematically equivalent to a frequency multiplier with tandem connections. Their performance can be related to that of the latter, by the method of superposition in terms of sequence quantities. As in symmetrical component analysis sequence impedances will be used to represent balanced sets of linear network elements. In this case however, sequence voltages and currents are represented in the frequency domain by infinite series of discrete harmonic coefficients. Hence the expression $V_{0q}(t) = I_{0q}(t)\, Z_q$ will imply

$$V_{0q}(t) = \sum_{n=0}^{\infty} I_0(j(q+np)\omega_0) Z_q(j(q+np)\omega_0) \epsilon^{j(q+np)\omega_0 t}$$

and Z_q must be indicated explicitly by a function of frequency $Z_q(j\omega)$. Fortunately, in practice most frequency multiplier sequence impedances can be closely approximated by series combinations of resistance and inductance, so that such a function can be characterized by a single complex number $Z_q(j\omega_0)$ or $Z_q(jq\omega_0)$ for instance.

If the nonlinear element terminals are considered as balanced p-phase source terminals superposition can be used to obtain sequence impedances in the linear network as seen from these elements. Sequence networks are obtained for a single nonlinear section in the form shown in Fig. 6. In this diagram $V_{0q}(t)$ refers to the external driving source, $Vn_{0q}(t)$ refers to the sequence components of the voltages at the nonlinear element terminals.

If there is an equal, unique relationship between time-varying terminal quantities for each member of the set of nonlinear elements, there must be a one-to-one relationship between sequence voltages and currents for the p-phase set when the prescribed symmetry assumptions are met by the entire circuit. When saturating cores with multiple windings are used, the nonlinear relationship exists between mutually induced voltages or flux and the total ampere turns acting on each core. By various winding connections sequence currents are allowed to flow in different paths, but the effect on the total ampere turns will be the same if external driving forces and sequence impedances are identical. The sequence network diagram can then be used to indicate the equivalence of various linear network connections and results obtained for one circuit in terms of sequence quantities are applicable to all equivalent versions.

For the reduction networks of Fig. 5 if impedance branches were used in each branch, or nodal reduction, it is apparent that the number of independent branch impedance functions is equal to the number of reduction stages. Similarly the number of independent sequence impedances and the number of isolated combinations of sequence quantities is equal to the number of reduction stages.

The tandem circuit can be used effectively as a model for quantitative analysis of equivalent connections as some simplification usually is obtained. Note however that physically, this model will not be equivalent in all cases, because of a restraint which it places on sequence impedances. For example in Fig. 5(A) the sequence impedances are equal to

$$Z_0 = Z_p + 6Z_L$$

$$Z_1 = Z_2 = Z_3 = Z_4 = Z_5 = Z_p$$

where Z_p is the series impedance in the first stage. Hence the nonlinear elements always see a zero-sequence impedance equal to or larger than that of other sequences. Using saturating cores with mutually coupled windings this constraint is not necessarily imposed. Mathematically however, there is no difference between the two circuits and identical sequence impedances can be obtained possibly requiring a fictitious, negative branch impedance in the model. A similar situation occurs commonly, when using a T equivalent circuit for a 2-winding transformer.

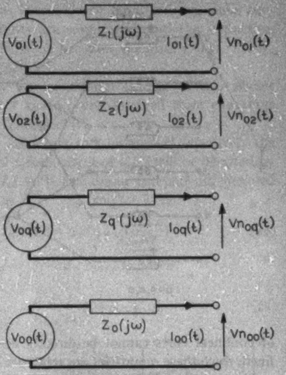


Fig. 6. Sequence networks as seen from a single set of nonlinear elements

In finding equivalent circuits one practical circuit variation deserves some further consideration. When the number of independent nonlinear elements in a single section are even, a normal p-phase driving source, if it contains only odd harmonic quantities can be obtained from $p/2$ of the phase voltages, by the use of simple transformers.

From equation 11 it is seen that for q odd

$$V_{p/2q}(t) = \epsilon^{jq\pi} V_{0q}(t)$$

$$= -V_{0q}(t)$$

Hence the odd harmonic voltages in phase $k+p/2$ are reflections of those in phase k. Rather than use transformers on the set of $p/2$ source voltages, phases $k+p/2$ can be driven directly from phase k. If a zero sequence source such as a d-c current is applied externally it must also be reversed on phases $k+p/2$. Other even sequence components generated in these phases will also be reversed so that various changes are necessary in phase-reducing networks, in order to maintain symmetrical frequency multiplier operation. Examples of such modified circuits are shown in Fig. 1(D) and in reference 3. The sequence impedances can be found for these circuits by superposition, applying modified phase, sequence components at the nonlinear element terminals.

Quantitative Circuit Analysis

The steady-state relationship between sequence voltage and currents on a set

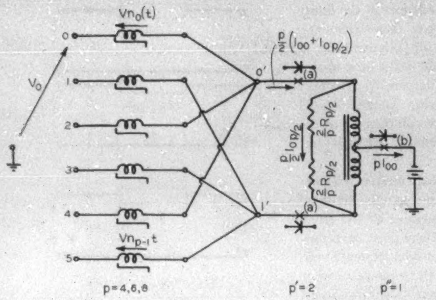


Fig. 7 (above). Example of a 6–2–1 reduced magnetic frequency multiplier

$R_{p/2} = R_{\text{load/phase}}$
$Z_{q=0}$ for $q \neq p/2$

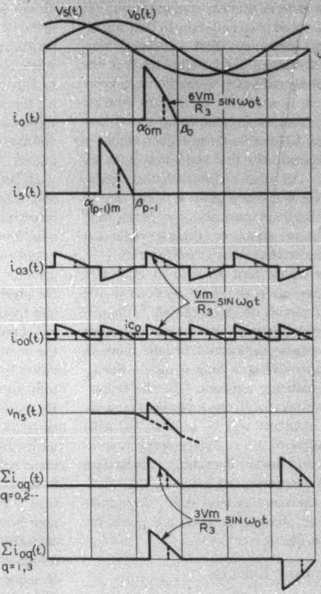


Fig. 8 (right). Waveforms for the 6–2–1 reduced frequency multiplier

of nonlinear cores cannot be directly defined, since these quantities are related to total element voltages and currents by linear superposition. Time domain relationships are defined however, and using these on the total voltage and currents appearing in a tandem frequency multiplier will give a quantitative solution in terms of sequence components. These will be obtained directly in time functional form and can then be characterized by rms, average or harmonic functional values.

SATURATING CORE FREQUENCY MULTIPLIERS

For the analysis of these saturating core circuits a piecewise linear approximation of core **B-H** loops has been found suitably accurate. Magnetic frequency multipliers operating under load provide large enough magnetic intensities on cores to render the effect of saturated reactance essentially linear. However the combined effect of this with leakage reactance and coil resistance must usually be considered to obtain accurate analytical results. Time response of the circuits under cyclic switching transients have been obtained using the Laplace transformation.[1] The equations and results in such an analysis are found by using various *s* functional impedances. From the previous study it is seen that if these impedances are related to sequence impedances the analysis will apply to a number of linearly equivalent circuits. The use of equivalence in this manner is of great advantage since the complete solution for one circuit in terms of two or three independent parameters has been found to require considerable time and effort even with the aid of a digital computer.

In order to illustrate the use of generalizations obtained in this discussion and also to indicate an important use of diodes in tandem sections of a frequency multiplier, a highly idealized multiplier circuit will now be treated quantitatively. Such an analysis has been used effectively to make comparisons between frequency multipliers on the basis of optimum performance.[2,3] The circuit chosen here uses the 6–2–1 reduction network of Fig. 5(C) and is a d-c controlled frequency multiplier.

The idealizing assumptions which are made in this analysis are:

1. The nonlinear cores are ideal having infinite unsaturated and zero saturated state reactances. Hysteresis and eddy currents are neglected.
2. Excitation is provided by an ideal, balanced, sinusoidal voltage, driving source and an ideal d-c control source. Saturating reactors are designed to obtain knee-point flux excursions, when no control current flows.
3. Windings have zero resistance and linear transformers and diodes are ideal.

In order to rate various equivalent connections it is necessary to find the rms value of isolated groups of sequence currents. For the 6–2–1 reduction all sequences currents flow in the first section, sequences $q=0$, $p/2$ flow in the second ($p'=2$) section and the zero sequence in the last. Where the circuit is inverted to operate from a 3-phase source the combined rms value of odd sequences also must be known for power-factor determination. The design of the ideal branch reducing transformer will be determined by the maximum volt-second integral of the third-sequence voltage across the load and the rms value of zero sequence current. For comparison purposes, the size of the saturating reactors can be obtained by determining their combined volt-amperage rating as a 2-winding power transformer.[2,3] A similar "rating factor" is applied to the ideal transformer. Capacitive power factor correction at the source will require knowledge of the reactive component of fundamental frequency input current. The modulus of the first output component $I_{0p/2}(jp/2\omega_0)$ will also be obtained.

In a piecewise linear analysis of this type it is found that similar modes of operation occur in circuits of different original phase order p, which have the same reduced phase sections, i.e., 6–2–1, 4–2–1, 8–2–1. In obtaining results from one analysis if the factor p is used explicitly, numerical results although different, can be directly obtained for all such circuits. The same is true in analyses which include internal reactance effects. The circuit of Fig. 7 for $p=6$ is treated in this manner.

It is noted in Fig. 7 that the d-c control source is placed in series in the zero sequence path of the tandem circuit. Since d-c power is not necessary for the operation of ideal frequency multipliers, no d-c resistance is placed in this path. Hence any desired direct current is obtained with an infinitesimal direct control voltage. The circuit operation however is entirely equivalent when a parallel zero sequence path with current limiting resistance is provided, along with a shorted final branch. Usually this parallel path will use mutually coupled windings on the saturating core. Looking into the linear circuit from the cores gives the same sequence impedances, and for a given direct control magnetomotive force, operation is identical. Chokes are usually necessary in frequency multiplier circuits to allow a low resistance d-c path in the zero sequence, along with a high harmonic-frequency impedance.[1-3]

Using the time symmetry condition all steady-state sequence quantities can be found if node and branch time functional quantities can be defined for one subperiod $T_{0/p}$. This period is conveniently taken from the angular time α_0 when core N_0 saturates.

For

$$\alpha_0 \leq \omega_0 t < \alpha_0 + \frac{2\pi}{p}$$

with N_0 saturated $2V_0$ appears across the ideal transformer and

$$p/2 i_{0p/2}(t) = \frac{2V_0(t)}{\frac{4}{p} R_{p/2}}$$

$$i_{0p/2}(t) = \frac{V_0(t)}{R_{p/2}} \tag{16}$$

For

$$V_0(t) = V_M \sin \omega_0 t$$

$$i_{0p/2}(t) = \frac{V_M}{R_{p/2}} \sin \omega_0 t \tag{17}$$

Since no net ampere-turns act on the ideal transformer

$$p/2 i_{00}(t) = p/2 i_{0p/2}(t)$$

$$i_{00}(t) = \frac{V_M}{R_{p/2}} \sin \omega_0 t \tag{18}$$

$$i_0(t) = p/2\left[i_{00}(t) + i_{0p/2}(t)\right] = p\,\frac{V_M}{R_{p/2}} \sin \omega_0 t \tag{19}$$

$$i_k(t) = 0 \text{ for } k \neq 0 \tag{20}$$

Core N_0 comes out of saturation at $\omega_0 t = \beta_0$ when $i_0(t) = 0$ during this interval. Hence β_0 is defined by $\sin \beta_0 = 0$

$$\beta_0 = \pi \tag{21}$$

For

$$\beta_0 \leq \omega_0 t < \alpha_0 + 2\pi/p \text{ all } i_{kq}(t) = 0$$

For this mode of operation, termed mode I, only one core is saturated at any given time. Equations 17–21 will define all isolated sequence components for this mode.

Mode II will occur if at some time in the period, two cores are simultaneously saturated. The initiation of mode II is indicated by the flux in the remaining cores N_k for $k \neq 0$. In the period $\alpha_0 < \omega_0 t < \alpha_0 + 2\pi/_p$, Core N_{p-1} will be the one to initiate mode II since its flux is at the saturation value at $\beta_{p-1} = \beta - 2\pi/p$. The voltage across N_{p-1} is shown in Fig. 8. For the period $\beta_{p-1} \leq \omega_0 t < \alpha_0$

$$vn_{(p-1)}(t) = V_M \sin(\omega_0 t + 2\pi/p)$$

For the period $\alpha_0 \leq \omega_0 t < \beta_0$

$$vn_{(p-1)}(t) = V_{p-1}(t) + i_{0\,p/2} R_{p/2}$$

and using equation 17

$$vn_{(p-1)}(t) = V_M[\sin(\omega_0 t + 2\pi/p) + \sin \omega_0 t)] \tag{22}$$

$$\varphi n_{(p-1)}(t) = \frac{1}{\omega_0} \int_{\beta_{p-1}}^{\omega_0 t} Vn_{(p-1)}(t)\, d\omega_0 t + \varphi_s \tag{23}$$

From Fig. 8 it is seen that $\varphi n_{(p-1)}(t)$ will reach a minimum at α_0 and then again approach φ_s dependin on the value of α_0. At the limit of mode I

$\int_{\beta_0}^{\omega_0 t} vn_{(p-1)}(t)\, d\omega_0 t$ will reach a maximum positive value of zero at a time $\omega_0 t > \alpha_0$ given by

$$vn_{(p-1)}(t) = 0$$

From equation 22 this time is $\omega_0 t = \pi - \pi/p$.

The limiting firing angle for mode I, α_{0M}, is then obtained from equation 32. Since

$$\int_{\pi-2\pi/2}^{\alpha_{0M}} \sin(\omega_0 t + 2\pi/p)\, d\omega_0 t + \int_{\alpha_{0M}}^{\pi-\pi/p} (\sin \omega_0 t + \sin(\omega_0 t + 2\pi/p)\, d\omega_0 t = 0 \tag{24}$$

which gives

$$\alpha_{0M} = \cos^{-1}(1 - 2\cos \pi/p) \tag{25}$$

Operation in mode II during the interval in which N_0 and N_{p-1} are saturated can be found by inspection of Fig. 7. The voltage $v_0(t) + v_{p-1}(t)/2$ is applied to the zero sequence path and $v_0(t) - v_{p-1}(t)/2$ to the $p/2$ sequence path. Obviously this operating condition cannot be permitted in an ideal circuit of this type since $Z_0 = 0$. However in practical cases it can be used as it is found that with sufficient zero sequence reactance more power output is attainable. If $Z_0(j\omega)$ for $\omega \neq 0$, is allowed to become very large zero-sequence currents are suppressed and the reduction network becomes essentially that of Fig. 5(B).[2]

The maximum ratings of this circuit for $\alpha_0 = \alpha_{0M}$ are obtained from equations 17–19 and 25.

Since it can be applied separately the d-c component of $i_{00}(t)$ will be denoted by I_{c0} with remaining components denoted as $i_{00}'(t)$, $I_{00}'(\text{rms})$. Then

$$I_{00}(\text{rms}) = [I_{c0}^2 + I_{00}'(\text{rms})^2]^{1/2} \tag{26}$$

$$I_{c0} = p/2\pi \int_{\pi-2\pi/p}^{\pi} i_{00}(t)\, d\omega_0 t$$

$$= \frac{V_M}{R_{p/2}} \frac{p}{2\pi} (1 + \cos \alpha_{0M}) \tag{27}$$

$$I_{00}(\text{rms}) = \frac{V_M}{R_{p/2}} \left[p/2\pi \int_{\pi-2\pi/p}^{\pi} i_{00}^2(t)\, d\omega_0 t \right]^{1/2}$$

$$= \frac{V_M}{R_{p/2}} [p(\pi - \alpha_{0M} + \tfrac{1}{2} \sin 2\alpha_{0M}/4\pi)]^{1/2} \tag{28}$$

$I_{c0}'(\text{rms})$ is found from equations 26 and 27. Since $i_{0p}/_2(t) = \pm i_{00}(t)$ depending on the interval

$$I_{0p/2}(\text{rms}) = I_{00}(\text{rms}) \tag{29}$$

using equation 19 the value of $I_0(\text{rms})$ is related to $I_{00}(\text{rms})$ by

$$I_0(\text{rms}) = \sqrt{p}\, I_{00}(\text{rms}) \tag{30}$$

Fourier Analysis of $i_c(t)$ can be used to find the active and reactive fundamental frequency components. However the active component is simply related to the power output

$I_{0x}(j\omega_0)$ referring to the rms value of the reactive component

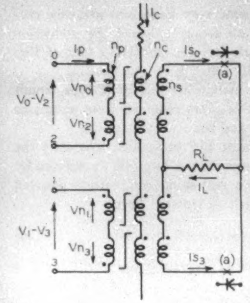
$$= -\frac{0.707}{4\pi} p \frac{V_M}{R_{p/2}} [1 - \cos 2\alpha_{0M}] \tag{31}$$

$$I_{0A}(j\omega_0) = \frac{p_{\text{out/phase}}}{0.707 V_M} = \frac{I_{0p/2}(\text{rms}^2) R_{p/2}}{0.707 V_M} \tag{32}$$

The complex value of the first harmonic output component is obtained by Fourier Analysis of $I_{0p/2}(t)$ expressed as an rms magnitude $I_0(jp/2\omega_0)$ becomes

$$\frac{V_M}{R_{p/2}} \frac{0.707}{\pi} \frac{4p}{p^2-4} \left\{ \left[(-1)^{p/2+1} + \frac{p-2}{4} \cos(p/2+1)\alpha_{0M} - \frac{p+2}{4} \cos(p/2-1)\alpha_{0M} \right] + j\left[\frac{p-2}{4} \sin(p/2+1)\alpha_{0M} - \frac{p+2}{4} \sin(p/2-1)\alpha_{0M} \right] \right\} \tag{33}$$

The rating factor[2] of the saturating cores is given by



$Vn_k(t)$ SEQUENCE COMPONENT PHASE ANGLES

k	Vn_{k1}	Vn_{k2}	Vn_{k3}	Vn_{k0}
0	→	→	→	→
1	↓	←	↑	→
2	←	→	←	→
3	↑	←	↓	→

Fig. 9 (above). Experimental circuit of the p–2–1 type, for p = 4

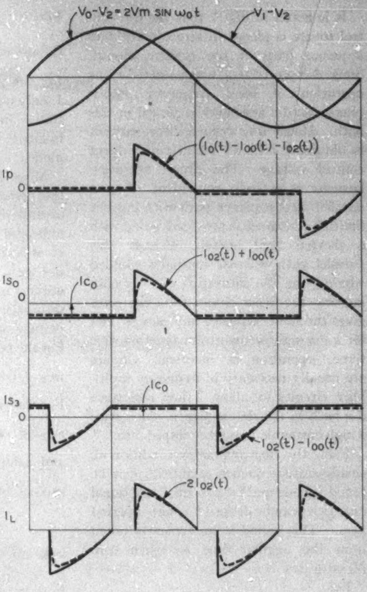


Fig. 10 (right). Predicted and observed waveforms for the experimental 2x multiplier

———Predicted results
- - - - -Experimental results

$$\text{RF (cores)}=V_M\frac{\frac{0.707}{2}I_0(\text{rms})}{p_{\text{out,phase}}} \qquad (34)$$

Note that with separate windings for various sequences this factor would be increased. If one separate winding were provided for zero sequence, and $p/2$ sequence quantities, the rms value of the remaining sequences in the primary winding (from equation 10) is given by

$$I_p(\text{rms})=[I_0{}^2(\text{rms})-I_{00}{}^2(\text{rms})-I_{0p/2}(\text{rms})]^{1/2}$$

The rating factor then becomes

$$V_M\times\frac{\frac{0.707}{2}I[_p(\text{rms})+\sqrt{I_{00}{}^2(\text{rms})+I_{0p/2}{}^2(\text{rms})}]}{p_{\text{out/phase}}}$$

on a primary turns basis.

A rating factor can also be obtained on the same output power basis for the ideal transformer. Its equivalent fundamental frequency voltage rating per phase is given by

$$\frac{1.11}{\pi}\int_{\pi-2\pi/p}^{\pi} i_{0p/2}R_{p/2}d\omega_0 t$$

$$=\frac{1.11}{\pi}V_M(1+\cos\alpha_{0M}) \qquad (35)$$

Hence

$$\text{RF (trans)}=\frac{\frac{1.11}{2\pi}V_M(1+\cos\alpha_{0M})I_{00}(\text{rms})}{p_{\text{out/phase}}} \qquad (36)$$

The method of obtaining the time function representing all combined odd sequences is shown in Fig. 8. The rms value of these is given by

$$\left[\sum_{q=1,3,5\ldots p-1} I_{0q}{}^2(\text{rms})\right]^{1/2}$$

and is seen to be equal to

$$\frac{1}{\sqrt{2}}I_0(\text{rms}) \qquad (37)$$

Using equations 26 to 37 design data for multipliers of this type with $p=12, 8, 6, 4$ are given in Table I. Results are normalized with $V_M=1.0$, $R_{p/2}=1.0$.

As an example, the operation of an experimental 4-core frequency multiplier shown in Fig. 9 will be predicted on an ideal basis. Data for this unit are as follows:

Cores. Deltamax toroidal no. *3T5690 D4* (Arnold)
$2^1/_2$ by $3^3/_4$ by $1^1/_2$ inches – 4 mill
$n_p=273$, $n_s=546$, $n_c=826$

$V_0-V_2=1.414\times120\sin\omega_0 t$
$V_1-V_3=1.414\times120\sin(\omega_0 t-\pi/2)$

$R_L=50$ ohms

Using the primary turns as a basis

$$V_0(t)=84.8\sin\omega_0 t$$

The relative directions of the sequence components of $Vn_k(t)$ are shown in Fig. 9. Applying these to the primaries a sequence at a time gives

$$i_p(t)=i_{01}(t)+i_{03}(t)$$

$$i_{s0}(t)=\frac{n_p}{n}[i_{02}(t)+i_{00}{}'(t)]$$

$$i_{s3}(t)=\frac{n_p}{n_s}[i_{02}(t)-i_{00}{}'(t)]$$

$$i_c=\frac{n_p}{n_c}I_{c0} \text{ where } \frac{n_p}{n_c}I_{c0}+\frac{n_p}{n_s}i_{00}{}'(t)=i_{00}(t)$$

$$i_L=\frac{2n_p}{n_s}i_{02}(t)$$

$$Z_2=R_L\left(\frac{N_s}{N_p}\right)^2=12.5 \text{ ohms}$$

From the equations of the preceding section the waveforms can be predicted using the relation

$$i_{01}(t)+i_{03}(t)=i_0(t)-i_{02}(t)-i_{00}(t)$$

These are shown on a primary turns basis in Fig. 10 along the observed values. Ratings are obtained using Table I.

$$I_p(\text{rms})=[I_0{}^2(\text{rms})-I_{02}{}^2(\text{rms})-I_{00}{}^2(\text{rms})]^{1/2}$$

$$=\frac{84.8}{12.5}[0.956^2-0.478^2-0.478^2]^{1/2}$$

$$=4.58A$$

$$I_s(\text{rms})=\frac{1}{2}[I_{02}{}^2(\text{rms})+I_{00}{}'(\text{rms})^2]^{1/2}$$

$$=\frac{1}{2}\frac{84.8}{12.5}[0.478^2+0.298^2]^{1/2}=1.91A$$

$$I_L(\text{rms})=I_{02}(\text{rms})=\frac{84.8}{12.5}0.478=3.24A$$

$$I_c=\frac{273}{826}I_{c0}=\frac{273}{826}\left(\frac{84.8}{12.5}\right)(0.373)=0.837A$$

Measured values are found at the

Table I. Ideal Ratings for p–2–1 Reduced Frequency Multiplier with $Z_{p/2}(j\omega)=R_{\text{load/phase}}$ and all other $Z_q=0$

p	4	6	8	12
Output Frequency $n=1,2\ldots$	$(2+n4)\omega_0$	$(3+n6)\omega_0$	$(4+n8)\omega_0$	$(6+n12)\omega_0$
$\alpha_0 M$, degrees	114.5	137.1	148.0	158.7
I_{c0}	0.373	0.256	0.193	0.130
I_{00}'(rms)	0.298	0.232	0.179	0.122
I_{00}(rms)	0.478	0.345	0.264	0.178
$I_{0p/2}$(rms)	0.478	0.345	0.264	0.178
$I_0(jp/2\omega_0)$	0.422	0.312	0.237	0.1656
I_0(rms)	0.956	0.845	0.747	0.617
$\left[\sum_{q=1,3,5,\ldots}^{p-1} I_{0q}^2(\text{rms})\right]^{1/2}$	0.675	0.597	0.528	0.436
$I_{0x}(j\omega_0)$	−0.341	−0.313	−0.253	−0.1783
$I_{0A}(j\omega_2)$	0.322	0.168	0.0985	0.0447
p_{out}/phase	0.228	0.119	0.0696	0.0316
Rating F (cores)	1.48	2.51	3.79	6.90
Rating F (trans)	0.217	0.1332	0.102	0.067

approximate limit of mode I, in Fig. 11(A) to be about 10% lower than predicted values, since primary and secondary winding impedances have been neglected. Predicted values would nevertheless be useful for approximate coil design.

The theoretical rating factor for the saturating cores of this configuration is given by

$$\frac{0.707}{2}\left\{\frac{[I_0^2-I_{02}^2-I_{00}^2]^{1/2}+[I_{02}^2+I_{00}'^2]^{1/2}+I_{c0}}{I_{02}^2}\right\}=2.49$$

This value is considerably higher than the combined value for the cores (1.48) and the transformer (0.22) of the tandem circuit. This circuit has the advantages however, of lower input power factor and separate windings for control and output quantities.

SELF-SATURATION

One disadvantage of d-c controlled frequency multipliers such as those described in this and in past articles,[1–8] is their need of substantial externally applied control magnetomotive force. This disadvantage is felt in power amplifier applications[1] since the control source must have considerable power output capacity to achieve rapid amplifier response. It is also felt in converter-power-supply applications where control current remains fixed at full output level regardless of load. As others have demonstrated[1–8] magnetic frequency multipliers behave essentially as a current source with regard to output unless at the maximum control level. Unfortunately with large control currents, removal of load does not cause any corresponding drop in rms input currents which become then mainly reactive. If the frequency multiplier is forced to supply its own d-c component, then as in the case of magnetic amplifiers, operation becomes very nearly that of a controlled voltage source with high gain, and removal of load also automatically reduces input currents to a negligible value.

In the analysis of the ideal tandem circuit of Fig. 7 it has been noted that in the steady state no finite direct voltage is necessary in the zero sequence path. Furthermore, since the cores cannot pass any negative load currents the currents in this path are always either zero or positive. Hence diodes could theoretically be instantaneously inserted in a symmetrical fashion at points A or point B of Fig. 7 without affecting circuit operation. In these locations the diodes are able to sustain direct magnetomotive force on all the cores, and the infinitesimal, direct voltage source could then be removed without affecting steady-state system behavior. In practice the cores will have finite unsaturated impedances and a very small quantity of externally applied direct current will be necessary to sustain a given level of operation. Also because of this magnetizing component, small negative currents are normally required in the p-phase section of the circuit so that insertion of diodes at this point could cause an entirely different and undesirable mode of operation.

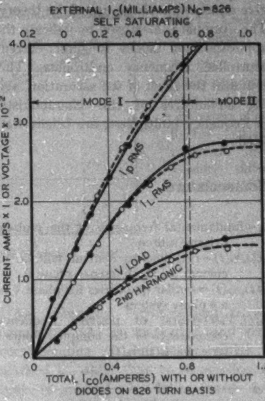


Fig. 11. Control characteristics for the 2x multiplier of Fig. 9 with and without self-saturation

——— Without self-saturation
– – – – With self-saturation

The same argument can be applied to any of the d-c controlled types of frequency multiplier. This has been proved experimentally with several basically different types such as typified by Fig. 5. It can be performed as indicated on equivalent connections such as those in Fig. 9 and also on circuits which are inverted to operate from lower phase sources. However, a strict requirement is provided in that the diodes must be placed symmetrically in the path of all zero-sequence components. In circuits having negligible zero sequence resistance no change is observed in the mode of operation, and results such as those of Table I can be used for rating purposes. With circuits having this property the steady state reverse voltage appearing on the diodes is usually only a small fraction of the line voltage, although with transients, peaks of this order do appear.

In Fig. 11 the control characteristics of the experimental unit of Fig. 9 are compared with and without self-saturation. Only the mode I type of operation has been included in the graphs although mode II operation is possible in both cases. The slight offset at the zero control point in Fig. 11(B) is caused by mismatches, or slight unbalances in the circuit. Changing the line voltage causes slight variations in this offset, however, the effect is not nearly as great as that observed in conventional magnetic amplifiers. The transient response of the device has been observed to be similar to that of a frequency multiplier without diodes operated at no load.[1]

Summary

A generalized steady-state theory of symmetrical frequency multiplier operation has been presented. This has provided principles for the synthesis of circuits and also a basis on which to obtain quantitative results applicable to various multipliers with equivalent connections. Considerable simplification in analysis is obtained by using a circuit model for a specific analysis.

A particular idealized frequency multiplier has been quantitatively analyzed to give an example of the use of the theory. The example is also used to indicate the application of diodes to direct-current controlled, frequency multipliers. This produces the effect of self saturation, and the advantages of this characteristic in frequency multipliers, are briefly outlined.

Nomenclature

ω_0 = fundamental frequency of the p-phase driving source
$i_0(t)$, $v_0(t)$ = real line to neutral voltage or current on the reference phase 0
$i_k(t)$, $v_k(t)$ = similar quantities for phase k or a p phase system
$I_k(t)$, $V_k(t)$ = total of positive frequency components of the complex Fourier series for $i_k(t)$, $v_k(t)$
$I_{0q}(t)$, $V_{0q}(t)$ = component functions of $I_k(t)$ $V_k(t)$ given by summation of the positive frequency harmonic components with $\omega=(q+np)\omega_0$
$i_{0q}(t)$, $v_{0q}(t)$ = real value of $I_{0q}(t)$, $V_{0q}(t)$
$I_{0q}(j\omega)$, $V_{0q}(j\omega)$ = complex harmonic spectrum of $i_{0q}(t)$, $v_{0q}(t)$ containing frequency components for $\omega=(q+np)\omega_0$
$I_{c0}=I_{00}(j0)$
I_{0q}, or I_{0q}(rms) = rms value of the real, time function $i_{0q}(t)$
Z_q or $R_q=Z_q(j\omega)$ linear driving source sequence impedance as seen from a p-phase set of nonlinear elements
I_{00}'(rms) = combined rms value of all zero-sequence components except the direct current
$I_{0x}(j\omega_0)$ = reactive component of fundamental frequency source currents
$I_{0A}(j\omega_0)$ = active component of fundamental frequency source currents
RF (cores) = rating factor of the cores
RF (trans) = rating factor of the center-tapped transformer or its volt-ampere rating as a fundamental frequency transformer
N_k = nonlinear element in phase k
$Vn_k(t)$ = total terminal voltage across N_k

References

1. TWO-PHASE MOTOR DRIVE USING STATIC FREQUENCY-DOUBLER AMPLIFIERS, G. W. Dick, P. P. Biringer, G. R. Slemon. *AIEE Transactions*, pt. I (*Communications and Electronics*), vol. 76, Nov. 1957, pp. 497–504.

2. MAGNETIC FREQUENCY MULTIPLIERS AND THEIR RATING. PART I—FREQUENCY TRIPLERS, William McMurray. *Ibid.*, vol. 75, Sept. 1956, pp. 384–90.

3. MAGNETIC FREQUENCY MULTIPLIERS AND THEIR RATING. PART II—FREQUENCY SEXTUPLERS, William McMurray. *Ibid.*, vol. 76, July 1957, pp. 289–93.

4. METHOD OF SYMMETRICAL CO-ORDINATES APPLIED TO THE SOLUTION OF POLYPHASE NETWORKS, C. L. Fortescue. *Ibid.*, vol. 37, pt. II, 1918, pp. 1027–1140.

Locking Range of an Oscillator for Different Nonlinearities

B. R. NAG
NONMEMBER AIEE

Synopsis: Dependence of the locking range of an oscillator on the character of the stabilizing nonlinearity is studied in this paper and it is shown that a lower order nonlinearity provides a wider range.

LOCKED oscillators are used in various fields of communication engineering for synchronizing frequencies of oscillators belonging to a common system. A desirable property of the oscillators is that the locking range for a fixed amplitude of the locking signal should be large. The locking range is mainly determined by the nature of the nonlinearity in the oscillator. In most of the theoretical studies, the oscillator nonlinearity is assumed to be cubic. It is the purpose of this paper to study the dependence of the locking range on the character of the nonlinearity. Theoretical derivations are based on van der Pol's method of approximation.[1] Experimental results for checking the theoretical derivations have been obtained with the help of an electronic differential analyzer.

Theoretical Derivation of Locking Range

The oscillator is considered to be described by a second-order equation which may be written as:

$$\ddot{x}-a\dot{x}+\mu f(\dot{x})+\omega_0^2\, x=E \sin \omega t \qquad (1)$$

where a is the initial negative damping coefficient, $\mu f(\dot{x})$ is the nonlinearity, and $E \sin\ \omega t$ is the locking signal.

The general solution of equation 1 consists of the forced oscillation of frequency ω, free oscillation of frequency ω_0 and the oscillations of harmonic and combination frequencies. If ω is close to ω_0, free oscillation is suppressed and the oscillator executes oscillation of forcing frequency only and is said to be locked. If a and μ are small, the harmonics of the forced oscillation have little effect and in the first order of approximation may be neglected. When study of the locked behavior of the oscillator is the objective, the solution is assumed to be of the form:

$$x=A \sin(\omega t+\phi) \qquad (2)$$

A and ϕ are determined by E, a, μ, ω and the functional form of $f(\dot{x})$. It is noted that since the damping coefficient is nonlinear, though the solution may be periodic in the steady state, in the transient conditions A and ϕ vary with time. But the variation is slow so that $\ddot{A}$, $\dot{A}\dot{\phi}$, $\ddot{\phi}$, $a\dot{A}$, $a\dot{\phi}$, $\mu\dot{A}$, and $\mu\dot{\phi}$ are negligible in comparison with A and ϕ.

Differentiating equation 2:

$$\dot{x}=\omega A \cos(\omega t+\phi)+\dot{A} \sin(\omega t+\phi)+A\dot{\phi} \cos(\omega t+\phi) \qquad (3)$$

Differentiating again and neglecting the small-order terms:

$$\ddot{x}=-\omega^2 A \sin(\omega t+\phi)+2\omega\dot{A} \cos(\omega t+\phi)-2\omega A\dot{\phi} \sin(\omega t+\phi) \qquad (4)$$

Substituting in equation 1 x, $\dot{x}$, and $\ddot{x}$ from equations 2, 3, and 4 respectively, and keeping only the terms of fundamental frequency contributed by $\mu f(\dot{x})$, one obtains:

$$(\omega_0^2-\omega^2)\, A \sin(\omega t+\phi) -\omega a A \cos(\omega t+\phi)+a_1 \cos(\omega t+\phi)+b_1 \sin(\omega t+\phi)+2\omega\dot{A} \cos(\omega t+\phi)-2\omega A\dot{\phi} \sin(\omega t+\phi)=E \sin(\omega t+\phi) \cos \phi-E \cos(\omega t+\phi) \sin \phi \qquad (5)$$

where a_1 and b_1 are the coefficients of $\cos(\omega t+\phi)$ and $\sin(\omega t+\phi)$ respectively, in the Fourier expansion for $\mu f[\omega A \cos(\omega t+\phi)]$.

On equating the coefficients of sine and cosine terms on the two sides of equation 5 one obtains:

$$2\omega\dot{A}=-E \sin \phi+\omega A a-a_1 \qquad (6A)$$

$$2\omega A\dot{\phi}=-E \cos \phi+[(\omega_0^2-\omega^2)/\omega]\, \omega A+b_1 \qquad (6B)$$

In the absence of the forcing signal, the amplitude of free oscillation is obtained from equation 6(A) on putting $\dot{A}=E=0$. Let the amplitude of free oscillation be called ρ_0. Also, at this stage, normalize the amplitude of the locking signal, detuning, and the amplitude of the locked

---

Paper 60-37, recommended by the AIEE Electronics Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., January 31–February 5, 1960. Manuscript submitted May 19, 1959; made available for printing November 20, 1959.

B. R. NAG is with the University of Calcutta, Calcutta, India.

The author is indebted to Prof. J. N. Bhar for his constant help and guidance.

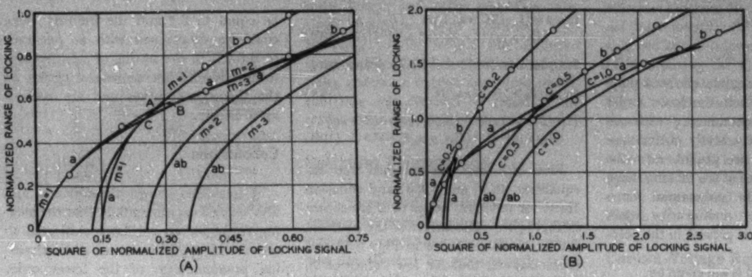


Fig. 1. Locking ranges

A—For polynomial-type nonlinearity, different values of m

B—For dead-zone-type nonlinearity, different values of C

oscillation, and call them F, δ and ρ, respectively. F, δ and ρ are defined as:

$$F=E/a\rho_0,\ \delta=(\omega_0^2-\omega^2)/a\omega,\ \rho=\omega A/\rho_0$$

Equations 6(A) and 6(B) in terms of these normalized quantities may be written as:

$$2\dot{\rho}=a[-F\ \sin\ \phi+\rho-a_1/a\rho_0] \tag{7A}$$

$$2\rho\dot{\phi}=a[-F\cos\ \phi+\delta\rho+b_1/a\rho_0] \tag{7B}$$

In the locked condition of the oscillator $\dot{\rho}=\dot{\phi}=0$. On putting $\dot{\rho}=\dot{\phi}=0$ in equations 7(A) and 7(B) and combining them, one obtains.

$$F^2=(\rho-a_1/a\rho_0)^2+(\delta\rho+b_1/a\rho_0)^2 \tag{8A}$$

$$\tan\ \phi=(\rho-a_1/a\rho_0)/(\delta\rho+b_1/a\rho_0) \tag{8B}$$

Equations 8(A) and 8(B) give, respectively, the amplitude and phase of the locked oscillation.

However, in order that the oscillator may remain locked, it is necessary that the equilibrium represented by $\dot{\rho}=\dot{\phi}=0$ should be stable. The stability is ensured if at the equilibrium values of ρ and ϕ (reference 2):

$$\frac{\partial\dot{\rho}}{\partial\rho}+\frac{\partial\dot{\phi}}{\partial\phi}<0 \tag{9A}$$

$$\frac{\partial\dot{\rho}\partial\dot{\phi}}{\partial\rho\partial\phi}\times-\frac{\partial\dot{\rho}}{\partial\phi}\times\frac{\partial\dot{\phi}}{\partial\rho}>0 \tag{9B}$$

In view of equations 7(A) and 7(B) and on using the equilibrium condition, inequalities 9(A) and 9(B) reduced to:

$$2-\frac{\partial}{\partial\rho}(a_1/a\rho_0)-a_1/a\rho\rho_0<0 \tag{10A}$$

$$\left[1-\frac{\partial}{\partial\rho}(a_1/a\rho_0)\right][1-a_1/a\rho\rho_0]+[\delta+b_1/a\rho\rho_0]$$
$$[\delta+b_1/a\rho\rho_0+\rho\frac{\partial}{\partial\rho}(b_1/a\rho\rho_0)]>0 \tag{10B}$$

For a fixed value of F, there is a maximum value of δ up to which the two inequalities are satisfied. This maximum value of δ, called δ (max), defines the range of detuning up to which the oscillator remains locked and is called the locking range of the oscillator. The locking range for a particular value of the locking signal may be obtained on reducing the two inequalities to two equations and solving each for δ simultaneously with equation 8(A). For some values of F, two values of δ (max) are obtained from inequality 10(B) in addition to a third one given by inequality 10(A). In most of these cases the largest of the three values gives the actual range of locking since the other two correspond to conditions of oscillation which are not permissible.[3,4]

The general equations deduced will now be used to determine the locking range for some special nonlinearities.

Locking Range for Polynomial-Type Nonlinearity

Let the nonlinearity be of the polynomial type given by:

$$\mu f(\dot{x})=k(\dot{x})^{2m+1}$$

In this case a_1 and b_1 are given by:

$$a_1=k(m)\ (\omega A)^{2m+1},\ b_1=0 \tag{11}$$

where $k(m)$ is a constant depending on m.

On substituting this value of a_1 in equation 7(A), ρ_0 is found to be given by:

$$\rho_0=[a/k(m)]^{1/2m}$$

Thus, a_1 may be written in terms of ρ as:

$$a_1=a\rho_0(\rho)^{2m+1}$$

On substituting for a_1 and b_1 in equations 8(A) and 8(B), ρ and ϕ in the locked condition are found to be given by:

$$F^2=\rho^2[(1-\rho^{2m})^2+\delta^2] \tag{12A}$$

$$\tan\phi=(1-\rho^{2m})/\delta \tag{12B}$$

Inequalities 10(A) and 10(B) also reduce to

$$\rho^{2m}>1/(m+1) \tag{13A}$$

$$(1-\rho^{2m})^2-2m\ (1-\rho^{2m})\ \rho^{2m}+\delta^2>0 \tag{13B}$$

The plot of δ (max) against F^2 for different values of m as obtained from the inequalities 13(A) and 13(B) in combination with equation 12(A) are shown in Fig. 1(A). The plot corresponding to inequality 13(A) is marked b and that corresponding to inequality 13(B) is marked a, while that corresponding to both is marked ab. It is noted that δ (max) as given by the plot marked a is larger than that given by the plot marked b or ab up to A (Fig. 1[A]). At A both the inequalities are just satisfied. The corresponding value of F^2 is obtained on solving simultaneously equation 12(A) and the equations derived from the inequalities and is found to be given by $(2m^2)/(m+1)^{(2+1/m)}$. The value of δ (max) as given by the plot marked a is slightly lower than that given by the plot marked b in the range between A and B (Fig. 1[A]). The value of F^2 corresponding to point B is obtained on considering that the two solutions for δ (max) as obtained from inequality 13(B) are identical at B. It is given by $(2m^2)(m+1)^2/(2m+1)^3$. In view of the discussion presented earlier, it follows that the locking range is given by the plot marked a for values of F^2 up to A and by the plot marked b for values of F^2 on the upper side of B. It should be noted that in the range extending between A and B, the locking range may be given by either of the plots depending on the direction of variation of δ.[3,4] However, the indeterminate zone is very small in extent and also the locking range as given by the two plots do not differ much.

On comparing the plots for different values of m in the light of the discussion, it is observed that the locking range is independent of the order of the nonlinearity for low values of F (normalized amplitude of the locking signal), while for higher values of F the lowest order nonlinearity gives the widest range.

Locking Range for Dead-Zone-Type Nonlinearity

In this section a dead-zone type of nonlinearity is considered, for which $f(\dot{x})$ may be written as:

$$\begin{aligned}\mu f(\dot{x})&=C(\dot{x}-E_b),\ \dot{x}>E_b\\&=0\quad -E_b<\dot{x}<E_b\\&=C(\dot{x}+E_b),\ \dot{x}<-E_b\end{aligned}\tag{14}$$

$\mu f(\dot{x})$ may be approximated by a polynomial in $\dot{x}$, in which the lower order terms increase in prominence with decrease in C. Hence a study of this nonlinearity provides an insight into the nature of dependence of the locking range on the order of the polynomial for a polynomial type of nonlinearity when more than one term is present in the polynomial.

In this case, a_1 and b_1 are given by:

$$a_1=\frac{2C}{\pi}\omega A\,[\cos^{-1}E_b/(\omega A)-(E_b/\omega A)\,\{1-(E_b/\omega A)^2\}^{1/2}]\tag{15A}$$

$$b_1=0\tag{15B}$$

The amplitude of free oscillation ρ_0 is given by:

$$\cos^{-1}E_b/\rho_0-(E_b/\rho_0)\,\{1-(E_b/\rho_0)^2\}^{1/2}=\pi a/2C\tag{16}$$

In terms of ρ and ρ_0, a_1 may be written as:

$$a_1=(2C/\pi)\rho\,\rho_0[\cos^{-1}(E_b/\rho\,\rho_0)-(E_b/\rho\,\rho_0)\,\{1-(E_b/\rho\,\rho_0)^2\}^{1/2}]\tag{17}$$

In this case, equations 8(A) and 8(B) reduce to:

$$F^2={}^2[1-(2C/\pi a)\,\{\cos^{-1}(E_b/\rho\,\rho_0)-(E_b/\rho\,\rho_0)\,(1-E_b^2/\rho^2\,\rho_0^2)^{1/2}\}]\tag{18A}$$

$$\tan\phi=[1-(2C/\pi a)\,\{\cos^{-1}(E_b/\rho\,\rho_0)-(E_b/\rho\,\rho_0)\,(1-E_b^2/\rho^2\,\rho_0^2)^{1/2}\}]/\delta\tag{18B}$$

Inequalities 10(A) and 10(B) reduce to:

$$(\rho\,\rho_0)/E_b>1/\cos(\pi a/2C)\tag{19A}$$

$$[1-(2C/\pi a)\cos^{-1}(E_b/\rho\,\rho_0)]^2-(2C/\pi a)^2(E_b/\rho\,\rho_0)^2\,[1-(E_b/\rho\,\rho_0)^2]+\delta^2>0\tag{19B}$$

Plots of δ (max) as obtained from inequalities 19(A) and 19(B) and equation 18(A) for different values of C are shown in Fig. 1(B). The nature of the plots are similar to those shown in Fig. 1(A) and it is observed that for low values of F, the locking range is almost independent of C while for higher values it increases with decrease in C. One may, therefore, conclude that even for a polynomial-type nonlinearity including more than one term, the locking range increases with greater predominance of the lower order terms.

Experimental Verification

The locking range for different nonlinearities considered has been obtained with the help of a differential analyzer. The limiting value of δ up to which the oscillation is locked was determined by observing the oscillator output on an oscilloscope. Appearance of the amplitude variation indicated unlocked condition. The initial damping was chosen to be equal to 0.1 and the limiting value could be determined with an accuracy of ±0.001 in the setting of ω_0^2. Experimental results are shown by circles on the theoretical curves and are seen to agree closely.

Conclusions

It has been shown theoretically and also verified experimentally that the locking range for an oscillator subjected to an external signal increases with increasing predominance of the lower order terms in the nonlinearity. For small amplitude signals, however, the locking range is not much influenced by the character of the nonlinearity.

References

1. Forced Oscillations in a Circuit with Non-linear Resistance, B. van der Pol. *The Philosophical Magazine*, London, England vol. 3, 1927, pp. 65–80.
2. Problème Général de la Stabilité du Movement (book), A. Liapounoff. Reprinted in "Annals of Mathematics Studies, Vol. 17," Princeton University Press, Princeton, N. J., 1947.
3. Forced Oscillations in Nearly Sinusoidal Systems, M. L. Cartwright. *Journal*, Institution of Electrical Engineers, London, England, vol. 95, pt. III, 1948, pp. 88–96.
4. On the Transformation of Singularities and Limit Cycles of the Variational Equations of van der Pol, A. W. Gillies. *Quarterly Journal of Mechanics and Applied Mathematics*, London, England, vol. 8, pt. 2, 1954, pp. 152–67.

Vol.
Pt. 1
1960

Current–Voltage Relationships in Negative Pulsed Corona

J. B. THOMAS
MEMBER AIEE

T. R. WILLIAMS
NONMEMBER AIEE

SOME INTERESTING properties of negative pulse corona have been noted by several investigators.[1,2] Among these properties are sparkover voltages and pulse currents considerably above d-c values. Such properties are desirable for those applications, such as electrostatic precipitation, where maximum corona loss may be sought. A more detailed study of negative pulse corona is therefore pertinent.

This paper presents the current–voltage relationships of negative pulse corona for a wide range of pulse lengths and repetition rates and for various values of load capacitance. In all cases a fixed cylindrical electrode structure was employed.

Apparatus and Procedure

A schematic diagram of the experimental equipment is shown in Fig. 1. The corona fixture consisted of a vertical aluminum discharge tube 12 in. (inches) long and 0.5 in. inside diameter and a coaxial nichrome wire 14 in. long and 0.010 in. in diameter. The wire was held taut by weights which were as heavy as the elastic limit of the wire permitted. The tube and wire were supported by insulators appropriately arranged to allow accurate centering of the wire in the coaxial tube. The ends of the tube were flared to prevent field concentrations.

The average current flow to the discharge tube was measured by a d-c milliammeter connected between the tube and ground as shown. Both the average and peak voltage across the corona gap were measured with appropriate meters.

The pulse generator was a conventional hard tube pulser with a pulse length continuously variable from 1 μsec (microsecond) to 50 μsec. Typical output waveforms for these two extreme cases are shown in Fig. 2(A) and Fig. 2(B). The pulse repetition rate could be varied from as low a value as desired to a rate corresponding to a duty cycle of approximately 0.25.

Paper 60-102, recommended by the AIEE Electronics Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., January 31–February 5, 1960. Manuscript submitted November 4, 1959; made available for printing December 9, 1959.

J. B. Thomas and T. R. Williams are with Princeton University, Princeton, N. J.

The authors wish to express their appreciation to Koppers Company, Inc., for the generous support which made this investigation possible. In particular they are indebted to J. W. Drenning for his interest, aid, and advice.

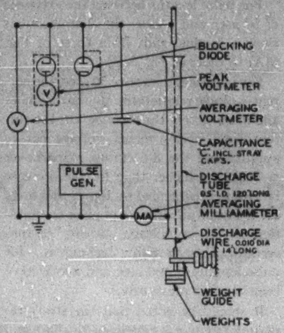

BLOCKING DIODE
PEAK VOLTMETER
AVERAGING VOLTMETER
CAPACITANCE "C" INCL. STRAY CAPS.
PULSE GEN.
DISCHARGE TUBE 0.5" I.D. 12.0" LONG
AVERAGING MILLIAMMETER
MA
V
V
DISCHARGE WIRE 0.010" DIA 14" LONG
WEIGHT GUIDE
WEIGHTS

Fig. 1. Diagram of apparatus

Measurements were made both with and without a series blocking diode between the pulse generator and the corona cell. When the diode was not present, the corona gap voltage was the pulse generator output voltage as in Fig. 2(A) and Fig. 2(B). With the diode in place, the gap voltage waveform was a function of the capacitance C between the discharge wire and ground. Typical waveforms are shown in Fig. 2(C) and Fig. 2(D). The capacitance C had an irreducible minimum due to stray capacitance and that of the measuring equipment but was varied above this minimum by adding an adjustable capacitor connected from the discharge wire to ground as shown. The total capacitance C was one of the parameters in the current–voltage relationships.

The usual precautions were taken to condition the discharge wire and obtain reproducible data. Since the corona characteristics were quite sensitive to pulse length and pulse repetition rate, these quantities were monitored continuously. Pulse length was adjusted from oscilloscope observations of the pulse generator output and pulse repetition rate was checked by counting the pulses on an electronic counter.

Sparkover Characteristics

As mentioned previously, sparkover voltages for negative pulse corona are considerably higher than negative d-c values and are affected by pulse length, pulse repetition rate, and load capacitance. Some typical data are shown in Fig. 3. Here sparkover voltages are shown for two duty cycles, 0.01 and 0.1, for three load conditions. For convenience the d-c sparkover value is shown as a horizontal line.

In all cases, as load capacitance increases, sparkover voltage decreases as the d-c case is more closely approximated. Also, as pulse length decreases for a constant duty cycle, sparkover voltage tends to decrease as the load capacitance becomes a more effective filter.

Another interesting quantity is the product $V_s\,I_m$, where V_s is the sparkover voltage and I_m is the average d-c current obtained just before sparkover. This product is a measure of the maximum obtainable corona loss. It is convenient to plot this quantity as is done in Fig. 4, where $V_s\,I_m$ has been plotted versus the parameter T_o/T_D. Here T_o is the off time or time betweeen pulses and T_D is a nominal fall time obtained as follows. The corona cell was assumed to have a constant resistance R of one megohm, and T_D was taken to be the product of R and the load capacitance C. For the case where no blocking diode was employed, 30 $\mu\mu f$ (micromicrofarad) has been arbitrarily assigned as the value of C for the purpose of normalizing the data so that it falls approximately along a single curve as in Fig. 4. As the ratio T_o/T_D decreases, the product $V_s\,I_m$ increases as the d-c case is approached. It should be noted that the d-c value of $V_s\,I_m$ is 81.4 volt-amperes and appears to be the asymptotic value of the curve represented by the points of Fig. 4.

Current-Voltage Relationships

The current-voltage characteristics of the pulse corona are difficult to present concisely since there are three major parameters affecting these characteristics: pulse length, pulse repetition rate, and load capacitance. However, Figs. 5(A) and 5(B) show some of the most interesting features of these characteristics.

If the pulse length is long, the data can be approximately normalized for each load condition. This situation is shown in Fig. 5(A) where the current–voltage curves for 50-μsec pulses have been plotted. Two load conditions are presented: the case where no blocking diode is employed, and the case where the load capacitance is approximately 383 $\mu\mu f$. For each of these two cases, data are given for four duty cycles. For convenience in comparison, the negative d-c characteristic has been added to the figure. It should be noted that the ordinate is not average current but average current divided by duty cycle. The abscissa is the height of the voltage pulses. For low load capacitance, the data for different duty cycles normalize along a single curve which is approximately the d-c characteristic ex-

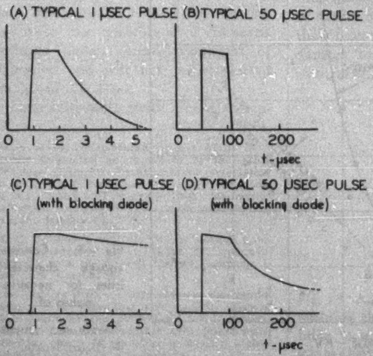


Fig. 2 (left). Typical voltage waveforms

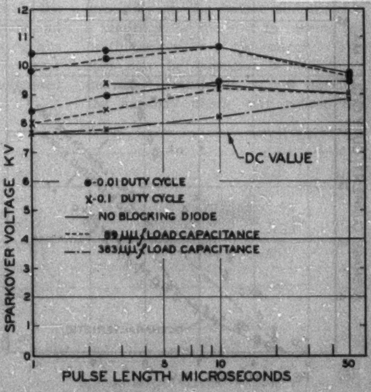


Fig. 3 (right). A comparison of negative pulse and d-c sparkover values

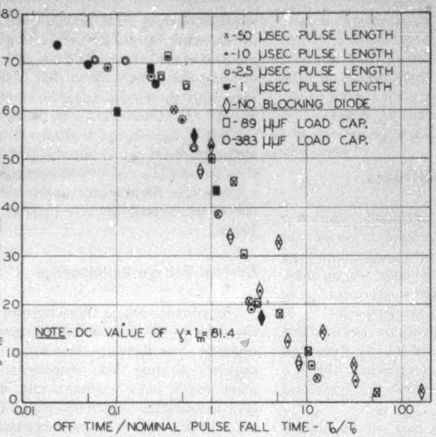


Fig. 4. Negative pulse corona power loss

cept that the range has been extended as a result of the higher pulse sparkover voltages. As the load capacitance increases the curve shifts and the normalization is not as complete, although the data for different duty cycles still lie approximately along the same curve.

For short pulse lengths, on the other hand, such normalization as described above is not possible. A typical situation is shown in Fig. 5(B) where data for 1-µsec pulses are presented. In this figure, average current is plotted directly versus peak voltage. The same two load conditions are considered as in Fig. 5(A). For the short pulses, the time between pulses becomes small as the duty cycle is increased so that a given load capacitance acts as a more effective filter. Consequently, for high duty cycles and high load capacitance, the situation approaches the d-c case as shown.

For pulse lengths between the extremes of 50 µsec and 1 µsec, the characteristics show a gradual shift from those of Fig. 5(A) to those of Fig. 5(B).

For pulse lengths above 50 µsec, the current–voltage relationship approaches more and more closely the d-c case, i.e. the current/duty cycle versus voltage curves approximate the d-c curves. Sparkover voltages are higher than for the d-c case however, as suggested by Fig. 3, and there are the usual advantages of pulse energization as far as voltage stability is concerned. Energization techniques for pulse lengths of 100 µsec and above have been studied previously.[3]

It should be noted that the study reported herein was restricted to short gap (1/2 in.) corona fields. For the larger electrode spacings of 3 to 6 inches encountered in some applications, further investigation is indicated.

Conclusions

With negative pulse energization it is possible to obtain corona losses of the same order as for the d-c case. This can be accomplished even with relatively low duty cycles. At the same time peak voltages considerably in excess of the d-c sparkover value can be maintained.

In addition to the above, pulse energization permits the selection of a wide range

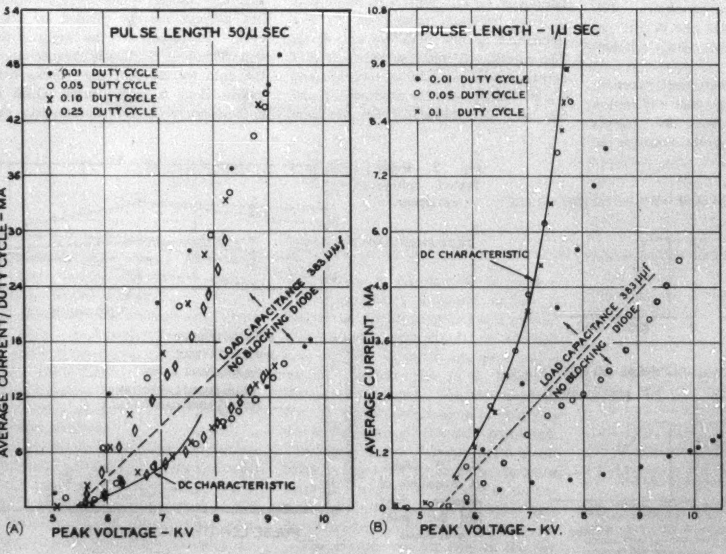


Fig. 5. Current-voltage characteristics for negative pulses of:

A—50-µsec length
B—1-µsec length

of current–voltage characteristics for a given corona cell electrode structure. Characteristics can be changed by varying pulse length, pulse repetition rate, and load capacitance.

References

1. Notes on Impulse Corona Studies in Air, H. J. Hall. *Physical Review*, New York, N. Y., vol. 72, 1947, p. 185.

2. Point-to-Plane Impulse Corona, D. B. Moore, W. N. English. *Journal of Applied Physics*, New York, N. Y., vol. 20, 1949, pp. 370–75.

3. A Pulse Method for Supplying High-Voltage Power for Electrostatic Precipitation, H. J. White. *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 71, Nov. 1952, pp. 326–30.

Voltage Transients Due to Arc Extinction

H. C. STEINER
FELLOW AIEE

R. W. STRECKER
ASSOCIATE MEMBER AIEE

HIGH TRANSIENT VOLTAGES may be developed in equipment associated with gas conduction tubes when there is a sudden cessation or change in the magnitude of the current that is being conducted through the tube. The conditions for the occurrence of this phenomenon have received considerable attention from those working in the field of gas conduction tubes as well as those who have applied such devices. Three conditions of operation of gas conduction devices have been recognized as contributing or being capable of creating the rapid rate of current change that is necessary to generate the high voltages observed: 1. cathode-spot extinction, 2. arc extinction, or gas "starvation" as it is sometimes called, and 3. the pinch effect. The pinch effect occurs when the current through the discharge reaches a sufficient magnitude to cause the conduction electrons to be pressed into a tight spiral because of the force of the self-induced magnetic field on the electron trajectory. This phenomenon is receiving considerable attention in the controlled thermal nuclear work that is currently being undertaken. Of the three, the pinch effect has not been demonstrated to be a contributing factor in the generation of transient voltages under normal conditions of operation and will not be considered further in this paper.

Langmuir, Mott-Smith, Tonks, Compton, and Fetz have studied and reported[1–5] on the physical factors involved in arc discharges at low pressures. Tonks, in a paper[6] presented before the Electrochemical Society in 1937, discussed the theory and phenomenon when a low pressure arc is "overloaded," and the resulting sudden change in current "creates surges of many thousands of volts." Hull,[7] in his development of hot cathode, grid-controlled, gas-filled-thyratron tubes, found the same high-voltage surges that had been found earlier in mercury-pool rectifying devices. Hull and Elder[8] as a result of their investigation of the phenomenon established certain quantitative data of permissible arc-current density as a function of gas density. Hansell[9, 10] and McArthur[11] have used the gas-starvation phenomenon as a nonlinear element for the generation of high-frequency currents.

It is the intent of this paper to review, briefly, the physical factors involved in the cathode-spot extinction and gas-starvation phenomena, to indicate the conditions under which such phenomena are likely to occur, to discuss the probable voltage generated, the factors within the tube which tend to limit the voltage and to outline some of the protective circuits that have been employed to prevent damage to transformers or associated equipment.

Arc Extinction Mechanisms

Cathode Spot Extinction

If one establishes an arc in a glass-enclosed mercury-pool tube that is operating at room temperature, one finds as the current is reduced that the arc becomes unstable and conduction ceases. At 3 amperes, for example, the arc seldom goes out. At 2 amperes, conduction ceases more or less frequently. At one ampere, it is very difficult to maintain the arc. If one substitutes metal for the glass wall of the mercury tube, the current for arc stability increases from the 3-ampere limit of the glass tube to 4 to 5 amperes for metal tubes of the same diameter. A further increase in current is required as the area of the container is increased. The explanation for the increased current was provided by Dr. K. H. Kingdon.[12] Fig. 1 shows the auxiliary anode current required for reasonable arc stability as a function of the operating temperature of a steel ignitron tube approximately 8 inches in diameter. The current to such an anode is composed of electrons which come from cathode spots on the mercury pool and the plasma, and from positive ion charges that are collected on the surface of the pool, the enclosing metallic envelope and any shields that are connected to the cathode. Free cathode spots on the mercury surface travel erratically on the surface and there is a tendency for the spot to burrow beneath the surface of the mercury. The spot may be quenched by collapsing mercury. Transiently, at low currents, the auxiliary anode may collect the

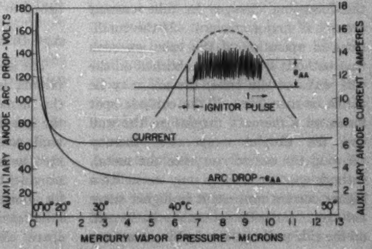


Fig. 1. Arc-drop voltage and minimum current for arc stability in a large metal ignitron as a function of temperature

---

Paper 60-76, recommended by the AIEE Electronics Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., January 31–February 5, 1960. Manuscript submitted February 6, 1959; made available for printing January 11, 1960.

H. C. Steiner is with the General Electric Company, Schenectady, N. Y.; and R. W. Strecker, formerly with the General Electric Company, is now at St. Charles, Ill.

The authors wish to acknowledge the careful reviews of the paper that were made by many of their associates and particularly the discussions with Mr. J. D. Cobine and Mr. A. O. Jensen which aided so greatly in the effort to develop a clear physical concept of the phenomena.

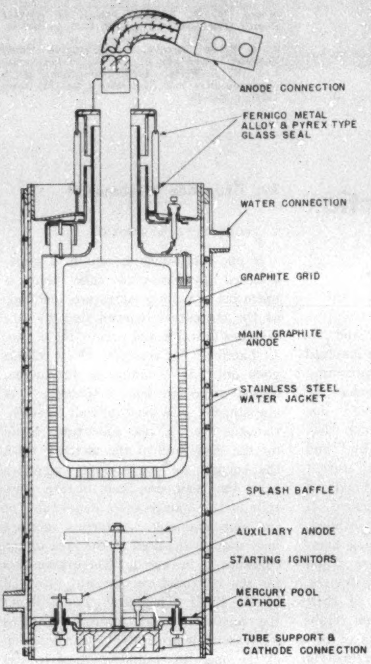


Fig. 2 (left). Cross section of a 400-ampere ignitron showing the general design

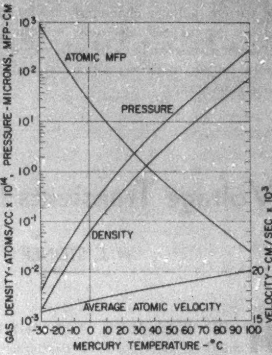


Fig. 3 (right). Properties of saturated mercury vapor[13,14]

required current from the plasma and this may extinguish the spot. Cathode-spot extinction is rapid, of the order of 10^{-8} seconds, and the auxiliary anode voltage would rise rapidly to high values if it were not possible to obtain a gradually decreasing supply of electrons from the decaying plasma or to re-establish new cathode spots. The arc-drop voltage under low temperature conditions is very oscillatory as indicated in the sketch of the observed waveshape. The oscillations disappear and the voltage becomes stable when the current is increased as with main anode conduction.

The circuit voltages generated from the foregoing changes in tube current are not of major concern. If the conditions of operation of the tube are such that cathode spots are established on the side walls of the tube, the arc may be stable or unstable. If the cathode spot forms on a mercury droplet on the wall and the droplet evaporates, thereby allowing the arc to run onto the metal, cathode-spot instability may occur at a higher current and result in higher surge voltages. Cathode spots may be formed on the side wall of the tube by extraneous magnetic fields forcing the arc into contact with the tube wall. Also, operation at very low temperatures and with the corresponding gas densities increases the arc drop and the probability of cathode-spot formation on the walls particularly if the tube is subjected to sudden high-current demands. Such spots may also form in hot cathode tubes if the available thermionic emission is insufficient to supply the current demand. In such cases, however, there is less likelihood of severe voltage surges developing because the mass of the parts in such tubes is usually small and the necessary temperature conditions are rapidly established to supply the required emission.

ARC STARVATION

The ability of arc to establish a normal plasma is dependent, as one factor, on the density of the available gas. As the gas density is lowered, the arc drop tends to increase. At very low pressures the gas density becomes insufficient to provide the molecules that are needed for ionization. The available ions are insufficient to neutralize the electron space charge and conduction changes rapidly from that of a gas discharge to that of a high-vacuum device. The effect is accentuated in tubes having grids or constrictions in the arc path. The current reduction as conduction changes from a gas discharge to high vacuum is very rapid, and transiently, voltages of the order of many thousands of volts may be generated in inductances associated with the current change. The general phenomenon has been termed "arc starvation."

Tests of an early model sealed ignitron tube will illustrate the characteristics of the phenomena. Fig. 2 shows the general design of the tube. The design employed a grid around the anode and the grid holes were of the order of a centimeter in diameter and a centimeter in length. When the tube was tested under low temperature conditions, rather violent voltage surges were found for cooling water temperatures in the range of 10–20 C (degrees centigrade). Below this range there was no conduction, while above it there was normal plasma conduction. The properties of saturated mercury vapor pertinent to this problem are shown in Fig. 3. Since the test loads were applied suddenly, the properties indicated correspond reasonably well to the conditions existing within the tube at the start of each test. At −20 C, which corresponds to a vapor pressure of 0.018 micron, the ignitor would fire satisfactorily, but the main anode, the grid and the small auxiliary holding anode which was approximately 2 inches from the ignitor would not conduct. As the temperature was increased to −10 C the main anode and grid still would not conduct but the holding anode would start. The discharge voltage trace was very oscillatory and the arc drop high, and the arc tended to go

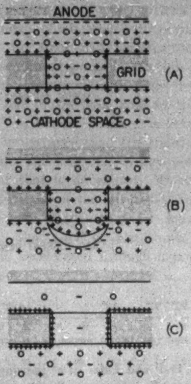


Fig. 4. Idealized sketch of the space conditions in the grid section of a gas tube showing gaseous atoms, ions and electrons in

A—Normal plasma
B—Plasma at reduced gas density in which a second sheath has formed
C—Transient high-vacuum conduction at very low gas densities

out frequently. At 3 C, the holding anode conducted and conduction started to the grid, but there was no measurable current conduction to the anode. At approximately 10 C, which corresponds to a vapor pressure of approximately 0.5 micron, conduction was established to the anode but conduction was erratic and voltage transients sufficient to break down protective gaps occurred. At 20 C, which corresponds to a vapor pressure of 1.2 microns, operation was stable over a current range of from a few amperes to several hundred amperes.

Although the general description of the phenomenon is straightforward, the physical changes involved are exceedingly complex. Figs. 4(A), (B), and (C) have been prepared to visually interpret three steps in the process of arc extinction. The cross sections show an enlarged view of one grid hole and the molecular, ion and electron entities that exist in the space. Normal conditions of conduction are shown in Fig. 4(A), the beginning of the surge conditions in (B), and the probable conditions at the end when the current has been reduced essentially to that of a high-vacuum tube in (C). These changes occur very rapidly, probably, as indicated later, in the order of 10^{-6} to 10^{-8} seconds.

Under normal conditions of arc conduction, Fig. 4(A), there are sufficient gas molecules in the space to provide the necessary positive ions for neutralization of any space charge because of the electrons conducted. The potential gradients in the plasma are low, probably of the order of a tenth of a volt per centimeter, and the ions drift slowly toward the cathode. Ions are lost to the grid and to the enclosing walls or enclosing shields if these are connected to the cathode. The plasma drop increases slightly to provide the additional ions needed to make up the losses. There will be a positive ion sheath around the grid if the grid is slightly negative with respect to the plasma and probably a very thin electron sheath at the anode. The thermal energy of the gaseous molecules is sufficient to maintain a reasonably uniform atomic distribution throughout the space, including the grid holes and the grid-to-anode region. The degree of ionization is low, probably of the order of a few per cent.

In Fig. 4(B) the gas density has been lowered to the point where insufficient positive ions are formed and the potential gradient through the plasma along the anode-to-cathode axis is increasing. As the arc drop increases to approximately 40 volts the energy of the electrons is increased and their efficiency of ionization increases. As the arc voltages increase above 40 volts, however, the efficiency of ionization decreases rapidly, and the beginning of the transient period is probably reached. At or prior to this time there is formed a double sheath[6] at the cathode side of the grid. There will be a concentration of positive ions on the anode side of this sheath and a concentration of electrons at the cathode side of the sheath as the arc attempts to establish conditions that will provide the needed supply of electrons. The drop through the plasma in the grid hole and in the grid-to-anode region further accelerates ions in the direction of the cathode and toward the grid surfaces. One further phenomenon occurs which probably accentuates the runaway conditions. As pointed out by Tonks, the electrons that are drawn across the cathode sheath provide a force of sufficient magnitude on the gas molecules within the grid hole to further deplete the density of the gas in that region. Very rapidly, then, the plasma disintegrates and the condition shown in Fig. 4(C) is obtained.

Fig. 4(C) is intended to represent the conditions of minimum current flow. At the instant of minimum current flow, positive ion sheaths cover the grid, the residual positive ions in the space on the cathode side of the grid have redistributed themselves so that the double sheath

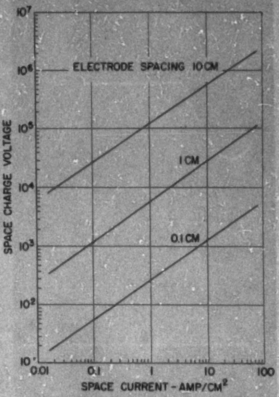


Fig. 5. Plot of the space-charge voltage between parallel plane electrodes, Childs-Langmuir equation[4]

$$J = 2.33 \times 10^{-6} \frac{V^{3/2}}{X^2} \text{ amperes/cm}^2$$

has disappeared, and the major part of the voltage that is applied between the anode and cathode exists between the grid and anode. Only a very small current would be conducted because of the high voltage drop that exists in high-vacuum conduction for relatively low current densities. Fig. 5 shows the calculated space charge for parallel plane electrodes separated by distances of the order of magnitudes that might be encountered in gas conduction tubes. These conditions exist only momentarily. Gas begins to flow back to re-establish equilibrium conditions. If the tube has a thermionic cathode, electron emission begins again, the plasma is re-established and the extinction cycle can begin again. This conduction-extinction-conduction cycle is shown in the oscilloscope traces of Fig. 6. These traces show the arc-drop voltage and current conducted in a xenon-filled thyratron operating with 1. normal gas pressure and 2. with the gas pressure reduced to near the surge limit. The tube was conducting as a half-wave rectifier from a 60-cycle source and the load current was resistance limited. Traces (A) and (B) of Fig. 6 show the arc-drop voltage and the anode current at the supply frequency. Trace (C) has been expanded to approximately 10 μsec (microseconds) per division to show the oscillations at the peak of the current wave for the low pressure condition and the absence of such oscillation for normal pressure. The extreme rapidity with

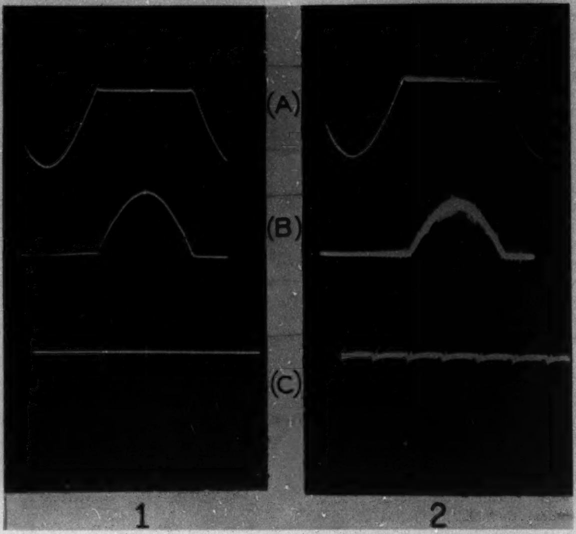


Fig. 6. Oscilloscope traces for 1. normal gas pressure and 2. low gas pressure in an inert-gas-filled thyratron showing

A—Arc voltage
B—Current at 60-cycle sweep frequency
C—Peak current at an expanded scale of 10 μsec per division

which the extinction and recovery takes place is indicated clearly in the expanded trace. If the device were a pool tube, it is probable that the cathode spots on the mercury pool have ceased and new cathode spots would have to be established in order to re-establish full current conduction through the tube.

The magnitude of the current that can be conducted through a grid as a function of mercury vapor pressure is indicated in Fig. 7. The data reproduced are from Figs. 6 and 7 of the Hull and Elder article.[8] The curves show the surge current for a large developmental thyratron and the corresponding current density through the grids for this and a smaller tube. Fig. 8 shows the remarkable increase in current capacity that occurs in an inert gas-filled thyratron when the gas density is increased just slightly above the critical surge value.

Voltage Generated

The voltage generated or, more properly, the voltage developed across the tube is, as indicated previously, directly proportional to the rate of change in current through circuit inductances and the tube. The magnitude of this voltage, if not restrained or limited, would be

$$e=-L\frac{di}{dt} \tag{1}$$

where

e = voltage developed
L = the inductance in henrys of the tube, the connecting leads, the leakage reactance of the transformer, and the reactance of the load circuit
$\frac{di}{dt}$ = the rate of current change in amperes per second, which for the purposes considered may be represented as finite changes ΔI and Δt

The voltage developed will be in such a direction as to attempt to maintain current flow through the tube, that is, the anode will become positive with respect to the cathode. The inductance L would be more properly represented by a transmission line since distributed capacities in the transformer and probably in the load circuit would tend to limit the L to only the end turns of the transformer windings. More or less of the load circuit would be involved, depending on the load wiring and the type of load. Primarily, however, the major factor is the rapid rate of current change and an evaluation of this will permit an evaluation of the voltage developed.

As indicated previously, there is probably a balance between the thermal diffusion of the gas into the grid-anode region and the drift velocity of the positive ions out of the region. As long as sufficient molecules exist in the grid holes to maintain a normal plasma, the plasma voltage gradient is relatively low. When there is a deficiency of gaseous molecules the voltage gradient increases and the velocity of the positive ions increases. With increased plasma voltage more ions are lost to the grid and to the double sheath when such sheaths are formed. One measure of the time of current extinction, therefore, would be the time that is required for an ion to travel from

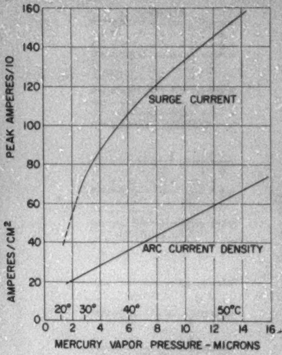


Fig. 7. Hull and Elder data[8] showing peak surge-current limit in a large thyratron and the maximum current carrying capacity for mercury vapor as function of pressure

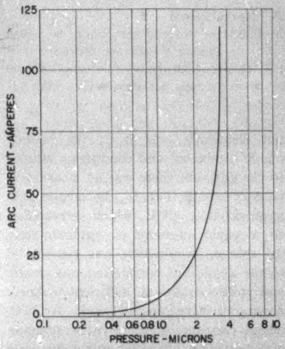


Fig. 8. Surge current limits in a xenon-filled thyratron showing the rapid increase in capacity as the critical gas pressure is exceeded

Vol.
Pt. 1
1960

the top of the grid to the bottom under the influence of whatever voltage gradient exists. For thick grids perhaps the more important factor would be the time an ion requires to traverse the distance from the center line of the grid hole to the grid wall. The actual distance is relatively of less importance than the speed of the ion under the influence of any potential gradient that is developed. If we consider the mass of the ion to be the same as the mass of the atom, the maximum velocity that such an ion could attain in being accelerated through a potential difference of V volts would be

$$1/2\, mv^2 = eV$$

where

m = mass of the ion in grams (3.33×10^{-22} for Hg)
v = velocity in cms per second
e = charge on the ion (4.767×10^{-10} esu)
V = potential difference in esu volts, or $V/300$ in practical volts

A mercury ion accelerated through a potential of one practical volt would attain a velocity of approximately 10^5 cm/sec (centimeters/second). A potential difference of 100 volts would increase this velocity 10 fold or to 10^6 cm/sec. These velocities would give times of the order of 10 μsec and 1 μsec respectively for a mercury ion to traverse a distance of 1 cm. Since the potential difference may be many fold the values discussed and the critical dimensions less than one centimeter, extinction times of the order of 0.1–1.0 μsec seem realistic. The fast oscilloscope trace, Fig. 6(C), of the current through a tube that is operating in the critical gas density range shows current cessations that are very rapid and of this order of magnitude.

As a first approximation of the voltage developed, we might consider only the commutating reactance of the transformers. For a 300-kw 275-volt rectifier, the commutating reactance per phase would be of the order of 10 microhenrys, and for a 5-kw 275-volt rectifier of the order of 1,500 microhenrys. If the entire inductance were involved in the transient and if there were no restraining forces, the voltage developed in the 300-kw unit for a current change of 10 amperes in a time interval of a quarter of a μsec would be approximately 400 volts. For a current change of 100 amperes and the same time interval, approximately 4,000 volts would be developed. For the 5-kw rectifier, current changes of one and 10 amperes in the same time interval would result in voltages of 6,000 and 60,000 respectively. These equations serve primarily to indicate that appreciable voltages may be developed. Quite obviously, the levels attained represent the unrestrained system and it is very doubtful that the actual inductance involved encompasses more than the first few turns of the transformer winding and a small part of the associated loads.

The results of tests made to determine the magnitude of the surge voltages "generated" indicate voltages within the range estimated. Generally, such tests are made by operating the tubes at very low gas pressures (10 C, for example, for mercury vapor tubes) and suddenly applying load. The test circuits usually include an air-core inductance in the anode lead to insure that the maximum voltage is developed. Suitable filters are placed across the power source and load. Voltages may be determined approximately from Lichtenberg figures[13] or more precisely from oscilloscope measurements. The maximum voltages developed under such conditions correspond rather well with the maximum voltage that a given design of tube will withstand. For example, it is relatively easy to design gas tubes for 10,000 volts, more difficult for 15–20,000 volts, and considerably more difficult in the 20–40,000 volt range. The highest test voltage measured—55 kv—within the authors' knowledge, was attained on a multigrid high voltage ignitron (*GL-6228/GL-506*) which has a rating of 20 kv forward and inverse and which withstands normally a potential test of approximately 50 kv. Voltages within the range of five to 25 kv are more representative of the majority of industrial tubes when operated under test conditions that would be expected to result in surges. The distribution of the magnitude of the surge voltages generated in a series of tests follows a normal distribution curve. Such a distribution is shown in Fig. 9 for two separate series of tests conducted on two types of tubes. If the calculated curve has been more precisely matched with the test data it would indicate that surge voltages approximately 20% greater than the maximum test voltage E found would be developed occasionally if a great many trials were run.

VOLTAGE LIMITS

There are several factors within the tubes themselves as well as in the external circuit that tend to limit the voltage developed. In the transformer, there is distributed capacity between the windings and ground and, in addition, there may be gaps, lightning arresters, or nonlinear resistors as protection for the transformer itself. From the tube standpoint, the conduction as a high-vacuum tube provides some slight limitation on the voltage that might be developed across the tube. More important is the likelihood of formation of new cathode spots on elements that are in the anode region. The ions under the conditions postulated have high energies and it becomes relatively easy to form new cathode spots. This is particularly true at the higher current levels where cathode spots may form on the inside of the grid and conduction then tends to become two arcs in series with the grid acting as anode with respect to the normal cathode of the tube and as cathode with respect to the anode. Finally, if there is no ionization present in the tube, there is still the probability of breakdown due to field emission. Experience indicates that very high vacuums must be attained and the tubes must be very "clean"[12] to sustain voltages in the 20–100-kv range.

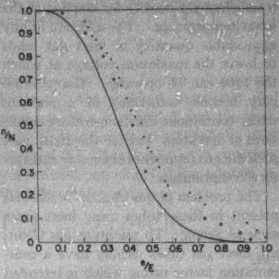


Fig. 9. Distribution of surge voltages in high-current low temperature (approximately 10 C) tests of two designs of mercury-vapor tubes compared with a Maxwellian distribution function[13]

—— Maxwellian distribution function

CONTROL METHODS

From the tube standpoint, the ratings and specifications are intended to define the range of operation without surging. Operation of the tube within the specified temperature limits for ambient or condensed mercury provides a high degree of insurance that voltage surges will not be developed for normal current demands or for fault currents. It is particularly important in the thermionic-cathode mercury-vapor tubes that normal vapor conditions be established since the vapor density changes so markedly with very small changes in temperature. Efforts to decrease the permissible lower operating temperature usually take the form of opening the grid structure or adding an inert gas (such as argon) to provide a sufficient gas density for starting at the lower temperature. The addition of any appreciable quantity of inert gas tends to lower the maximum voltage at which the tube can be operated. Gas cleanup may become something of a problem under continuous low-temperature conditions of operation, but for the most part such designs represent reasonably satisfactory compromises.

The problem of gas cleanup is likewise present in those tubes using inert gases such as xenon. To minimize gas cleanup, inert gas tubes usually have a commutation factor rating which is intended to permit full utilization of the tube capabilities. The commutation factor rating defines the circuit requirements in terms of the product of the rate of current decay (di/dt) at the end of the conduction cycle and the inverse voltage (de/dt) initially applied. High factors increase the rate of cleanup since the residual ions present in the space at the end of conduction attain higher energies with higher inverse voltages. These higher energy ions increase the sputtering and the rate of gas cleanup. Rapid gas cleanup may also occur near the end of the cathode life when the arc drop increases. A discussion of cleanup is beyond the scope of this paper except to recognize the phenomena and to indicate that satisfactory tube performance can be obtained by a combination of the design, materials, and processing and limiting the commutation factor. In general, other malfunctions tend to insure that the tube is removed from service before excessively high voltage surges develop.

From the equipment viewpoint, the application of gaps in conjunction with nonlinear resistors should provide adequate protection against the occasional possibility of high-voltage surges. Resistor-capacitor filters across the tubes or transformers should also provide effective attenuation. Such filters should be nonoscillatory with respect to the expected inductance. The capacitor should be capable of absorbing most of the energy stored in the inductance within its voltage rating. The initial surge voltage will be determined by the *IR* drop across the resistor.

Conclusions

This brief survey of the problems associated with arc extinction indicates that high transient voltages can be developed across vapor or gas-filled electron tubes when such tubes operate below their design gas densities. The sudden reduction in current, which creates the transient voltage, is dependent primarily on the gas density and tube geometry and not on the type of gas or vapor. Considerable testing has led to the conclusion that surging does not occur under normal gas densities or vapor temperatures. Gas cleanup in inert-gas tubes must be considered as an ultimate end of life possibility, but considerable field experience indicates that change in other characteristics is more likely to terminate the useful service period than is surging. Additional research and investigation are needed to clearly define these very interesting and complex phenomena.

The use of surge-voltage limiters should be considered, particularly for any essential service type of equipment. Without such limiters, voltage breakdown in the tube most probably determines the maximum voltage that can be developed. Standards for guidance would be desirable.

References

1. ELECTRICAL DISCHARGES IN GASES AT LOW PRESSURES, Irving Langmuir. *Journal*, Franklin Institute, Philadelphia, Pa., vol. 214, no. 3, Sept. 1932, pp. 275–98.

2. STUDIES OF ELECTRIC DISCHARGES IN GASES AT LOW PRESSURES, Irving Langmuir, Harold Mott-Smith, Jr. *General Electric Review*, Schenectady, N. Y., vol. XXVII, nos. 7, 8, 9, 11, July–Nov., 1924, pp. 449–55, 538–98, 616–23, 762–71.

3. GENERAL THEORY OF THE PLASMA OF AN ARC, Lewi Tonks, Irving Langmuir. *Physical Review*, New York, N. Y., vol. 34, no. 6, Sept. 15, 1929, pp. 876–922.

4. ELECTRICAL DISCHARGES OF GASES, Karl T. Compton, Irving Langmuir. *Review of Modern Physics*, New York, N. Y., vol. 2, pt. I, no. 2, Apr. 1930, pp. 123–242; vol. 3, pt. II, no. 2, Apr. 1931, pp. 191–257.

5. CONTROL OF A MERCURY ARC IN A VACUUM WITH A CONTROL GRID IN THE PLASMA, H. Fetz. *Annalen der Physik*, Leipsig, Germany, series 5, vol. 37, no. 1, Jan. 1940, pp. 1–40.

6. THEORY AND PHENOMENA OF HIGH CURRENT DENSITIES IN LOW PRESSURE ARCS, Lewi Tonks. *Transactions*, Electrochemical Society, Baltimore, Md., vol. LXXII, Seventy-Second General Meeting, Oct. 15, 1937, pp. 167–82.

7. HOT-CATHODE THYRATRONS, Albert W. Hull. *General Electric Review*, vol. 32, Mar.–Apr. 1929, pp. 213–23; vol. 32, no. 7, July 1929, pp. 390–99.

8. THE CAUSE OF HIGH VOLTAGE SURGES IN RECTIFIER CIRCUITS, Albert W. Hull, Frank R. Elder. *Journal of Applied Physics*, New York, N. Y., vol. 13, no. 6, June 1942, pp. 322–77.

9. MERCURY ARC OSCILLATOR, C. W. Hansell. *U. S. Patent no. 2,184,740*, Dec. 26, 1939.

10. ARC DISCHARGE CONTROL DEVICE, C. W. Hansell. *U. S. Patent no. 2,227,829*, Jan. 7, 1941.

11. GASEOUS ARC HIGH FREQUENCY GENERATOR, E. D. McArthur. *U. S. Patent no. 2,792,524*, May 14, 1947.

12. THE RELATION OF RESIDUAL IONIZATION TO ARC-BACK IN THYRATRONS, K. H. Kingdon, E. J. Lawton. *General Electric Review*, vol. 42, Nov. 1939, pp. 474–78.

13. GASEOUS CONDUCTORS—THEORY AND ENGINEERING APPLICATIONS (book), James Dillon Cobine. Dover Publications Inc., New York, N. Y., 1958.

14. FUNDAMENTALS OF ENGINEERING ELECTRONICS (book), William G. Dow. John Wiley & Sons, Inc., New York, N. Y., 1952.

Generalized Brillouin Flow

GORDON KENT
NONMEMBER AIEE

THE 1945 PAPER by Brillouin[1] presented an exact solution to the problem of space charge flow of electrons in a magnetic field. As is well known, this solution demonstrates the possibility of an electron beam of cylindrical or rectangular cross section in which the repelling forces due to the space charge are exactly balanced by the magnetic field forces so that the beam maintains its shape for an indefinite length. The many electron tubes which have been designed to make use of this type of flow as well as the large number of research papers published on problems associated with Brillouin flow constitutes substantial evidence of the impact on the electron tube industry of the Brillouin paper.

Although the many difficulties encountered in the use of Brillouin flow have in recent years made the less elegant but simpler method of confined flow-focusing more practical,[2] Brillouin's paper remains interesting as one of the early contributions to the cataloging of exact solutions to the electron flow problem. Since this problem is of such monumental difficulty that solutions of any degree of generality are elusive, the listing of exact solutions, starting with those of Langmuir and extending to the elegant contributions of Kirstein,[3–6] is equivalent to a series of snapshots from

---

Paper 60-31, recommended by the AIEE Basic Sciences Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., January 31–February 5, 1960. Manuscript submitted July 14, 1959; made available for printing November 18, 1959.

GORDON KENT is with Syracuse University, Syracuse, N. Y.

This work was done with the support of the National Science Foundation, research grant (*NSFG6314*).

which one may hope to acquire sufficient understanding to permit the development of solutions of increased generality. The solution presented here is of interest more as an addition to the list of exact solutions than as a result of great practical importance, although its implications for the focusing of beams of finite rectangular cross section are perhaps useful.

Many of the recent discoveries of new electron flow solutions are nascent in the approach recommended by Spangenberg in 1941.[7] In essence, Spangenberg makes use of a velocity potential to obtain a single partial differential equation representing the electromagnetic and the electrodynamic problems simultaneously. The Brillouin paper, in contrast, considers the two problems separately, seeking, ultimately, a set of self-consistent solutions. When the problem attacked by Brillouin is formulated along the line suggested by Spangenberg, the possibility of a somewhat general solution is immediately evident. As is shown subsequently, the Brillouin cylindrical and semi-infinite rectangular beams are special cases.

Derivation of Basic Equation

If it is assumed that the electrons originate with zero initial velocity from a cathode at which the normal component of the magnetic field vanishes, the generalized momentum is an irrotational field quantity so that

$$\mathbf{P}=\nabla W \tag{1}$$

It is also assumed, though not of necessity, that the velocity of the electrons is a single-valued function of the spacial co-ordinates and, further, that all field quantities are explicitly independent of time. In accordance with equation 1, the velocity is given by

$$\mathbf{v}=(1/m)\,(\nabla W+e\mathbf{A}) \tag{2}$$

where $\mathbf{A}$ is the vector potential from which the magnetic field is derived by the relation

$$\mathbf{B}=\nabla x\mathbf{A} \tag{3}$$

and e is the magnitude of the electron charge. With the velocity represented by equation 2, the energy equation takes the form

$$\Phi=\frac{1}{2em}\,(\Delta W+e\mathbf{A})^2 \tag{4}$$

where Φ is the electrostatic potential measured with respect to the cathode, and m is the electron mass. In accordance with equation 4, Poisson's equation becomes

$$\rho=(\epsilon_0/2em)\nabla^2\,[(\nabla W+e\mathbf{A})^2] \tag{5}$$

Finally, the continuity condition requires that the product of the left-hand sides of equations 2 and 5 be a solenoidal vector or, symbolically, that

$$\nabla\{(\Delta W+e\mathbf{A})^{\prime}\,[(\nabla W+e\mathbf{A})^2]\}=0 \tag{6}$$

In this derivation it has been assumed that $(v/c)^2<<1$ so that the magnetic field due to the electron motion is of no consequence and the applied magnetic field may be set independently of the motion. The vector $\mathbf{A}$ which appears in equation 6 is any vector consistent with equation 3, and to remove this ambiguity it will be assumed that $\mathbf{A}$ has zero divergence. With $\mathbf{A}$ determined, equation 6 is a fourth-order third-degree partial differential equation for the function W. Once W is determined, all other quantities of interest can be calculated from expressions 1 through 5.

Motion in Axial Magnetic Field

An axial (z-directed) magnetic field of strength B_0 is derivable from the solenoidal vector potential

$$\mathbf{A}=\frac{1}{2}\,B_0(-\hat{x}y+\hat{y}x) \tag{7}$$

where the circumflex indicates unit base vectors. There is little to be gained by substituting equation 7 in equation 6 to obtain a differential equation involving W and the co-ordinates only, because the resulting expression is cumbersome despite the simplifications that result from the fact that $\mathbf{A}$ is solenoidal and in the transverse plane and that

$$\nabla^2(e\mathbf{A})^2=e^2B_0^2 \tag{8}$$

Nevertheless, these properties of $\mathbf{A}$ lead directly to the observation that a function W which makes the charge density a constant, and at the same time is a solution of Laplace's equation, is also a solution of equation 6. Alternately, if the charge density depends on z only and W satisfies the further conditions that

$$\frac{\partial^2 W}{\partial_x{}^2}+\frac{\partial^2 W}{\partial_y{}^2}=0$$

$$\frac{\partial^2 W}{\partial_z{}^2}\,\rho+\frac{\partial W}{\partial_z}\,\frac{\partial\rho}{\partial_z}=0 \tag{9}$$

then W is a solution of equation 6 describing a more general motion.

To construct a solution satisfying equation 9, it is possible to put

$$W=m[\Omega xy+f[z]) \tag{10}$$

This assumption leads to the charge density

$$\rho=(\epsilon_0 m/2e)\left\{4(\Omega+\omega_L{}^2)+\frac{d^2}{dz^2}\left[\left(\frac{df}{dz}\right)^2\right]\right\} \tag{11}$$

where ω_L is the Larmor frequency. Evidently, the required functional dependence of ρ is provided and the first part of equation 9 is satisfied by the choice of W. The second part of equation 9 is satisfied providing $f(z)$ is a solution of the differential equation

$$\left\{4(\Omega^2+\omega_L{}^2)+\frac{d^2}{dz^2}\left[\left(\frac{df}{dz}\right)^2\right]\right\}\frac{df}{dz}=(2eJ_0/\epsilon_0 m) \tag{12}$$

where J_0, a constant, is the axial component of current density. Since a solution of the problem does not require the function $f(z)$ but only its first derivative, which is the z-component of the velocity, it is convenient to put $df/dz=s(z)$. In terms of the new variable, the first integral of equation 12 is

$$C_1=(\omega_L{}^2+\Omega^2)s^2+\frac{1}{2}\,s^2\left(\frac{ds}{dz}\right)^2-\left(\frac{eJ_0}{\epsilon_0 m}\right)s \tag{13}$$

Before obtaining the second integral of equation 12, it is enlightening to examine the significance of the integration constant C_1, and to classify the different kinds of solutions accordingly. The requirement that the term $s^2(ds/dz)^2$ be positive or zero in equation 13, leads to a greatest lower bound on C_1,

$$C_1\geqq-\frac{1}{4\omega^2}\left(\frac{eJ_0}{\epsilon_0 m}\right)^2=C_{\min} \tag{14}$$

and to greatest lower and least upper bounds on s,

$$\left.\begin{matrix}\dfrac{1}{2\omega^2}\left(\dfrac{eJ_0}{\epsilon_0 m}\right)\left[1-\sqrt{1+4\omega^2C_1/\left(\dfrac{eJ_0}{\epsilon_0 m}\right)^2}\,\right]\\ 0\\ \dfrac{1}{2\omega^2}\left(\dfrac{eJ_0}{\epsilon_0 m}\right)\left[1+\sqrt{1+4\omega^2C_1/\left(\dfrac{eJ_0}{\epsilon_0 m}\right)^2}\,\right]\end{matrix}\right\}\leqq s\leqq \tag{15}$$

where the greater of the two terms on the left of equation 15 applies. In the two previous expressions the substitution $\omega^2=(\omega_L{}^2+\Omega^2)$ has been made. In addition, a combination of equations 12 and 13 shows

$$\frac{d^2s}{dz^2}=-\frac{1}{s^3}\left[\left(\frac{eJ_0}{\epsilon_0 m}\right)s+2C_1\right] \tag{16}$$

Although equations 13 and 16 reveal that (ds/dz) and (d^2s/dz^2) become singular as s approaches zero, equation 12 remains satisfied, and s may assume the value zero. Such a point, however, cannot be a relative minimum.

It is now possible to classify solutions as follows:

1. If $C_1=0$, d^2s/dz^2 is negative so that s

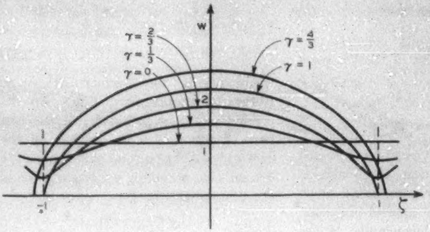


Fig. 1. Normalized velocity plotted against normalized distance with varying parameter γ

may have a relative maximum equal to the least upper bound in equation 15, and s may vanish.

2. If $C_{1\min}<C_1<0$, s may have relative minima and relative maxima given by equation 15, but s cannot vanish.

3. If $C_1=C_{1\min}$, s must be constant. In fact, $s=0$, or $s=eJ_0/2\epsilon_0 m\omega^2$ In the event that J_0 vanishes, still another solution is possible, but this is of no interest here.

In the foregoing discussion, only positive or zero values of s have been considered. Since the equations for s are reversible in z, all the remarks apply equally when s is a negative semidefinite function. The only difference in the analysis is the change of sign of J_0.

For solution 1 as well as solution 2 the point at which s is a maximum may be fixed at $z=0$ without loss of generlaity. The second integral of equation 12 then takes the form

$$|Z|=\frac{1}{\omega\sqrt{2}}\left\{\sqrt{(s_2-s)(s-s_1)}+\left(\frac{s_1+s_2}{2}\right)\times \sin^{-1}\frac{2\sqrt{(s_2-s)(s-s_1)}}{s_2-s_1}\right\} \qquad (17)$$

where

$$s_2;\ s_1=\frac{1}{2\omega^2}\left(\frac{eJ_0}{\epsilon_0 m}\right)\left[1\pm\sqrt{1+4\omega^2C_1\left(\frac{eJ_0}{\epsilon_0 m}\right)^2}\right]$$

For class 1, s vanishes when

$$Z=\pm\frac{1}{\omega\sqrt{2}}\left[\sqrt{-s_1s_2}+\frac{1}{2}(s_1+s_2)\times \sin^{-1}\left(\frac{2\sqrt{-s_1s_2}}{s_2-s_1}\right)\right] \qquad (18)$$

and for class 2, s takes the minimum value s_1 at

$$Z=\pm\frac{\pi}{\omega\sqrt{2}}\left(\frac{s_1+s_2}{2}\right) \qquad (19)$$

The ambiguity inherent in the appearance of the inverse sine function in expressions 17 and 18 is removed by the rule that for s in the interval $s_2=s={}^1/_2(s_1+s_2)$, the inverse sign takes values in $(0,{}^1/_2\pi)$, but for $s<{}^1/_2(s_1+s_2)$, the inverse sign is in $({}^1/_2\pi,\pi)$. In the case of class 2 solutions, s is a periodic function of z with period of $\pi/\pi\sqrt{2}(s_1+s_2)$, and with due caution all branches of the inverse function are applicable.

A graphical representation of the dependence of s on z is most readily obtained through the introduction of the following dimensionless variables:

$$w=(2\omega^2\epsilon_0 m/eJ_0)s$$

$$\zeta=(2\sqrt{2}\omega^2\epsilon_0 m/eJ_0)z$$

$$\gamma=\sqrt{1+(4\omega^2C_1)/(eJ_0/\epsilon_0\epsilon_0 m)^2}$$

In these terms, equation 17 becomes

$$\zeta=\frac{1}{\pi}\left[\sqrt{\gamma^2-(1-w)^2}+\sin^{-1}\sqrt{\gamma^2-(1-w)^2/\gamma)}\right] \qquad (20)$$

which evidently lends itself to the parametric representation

$$\pi\zeta=\theta+\gamma\sin\theta$$

$$w=1+\gamma\cos\theta \qquad (21)$$

Expression 21 describes a 1-parameter family of cycloids, as shown in Fig. 1. As C_1 increases from C_{1min}, γ increases from zero. When $\gamma<1$, the cycloids are curtate; and when $\gamma>1$, they are prolate. The ordinary cycloid results from putting $\gamma=1$, or equivalenty, $C_1=0$.

The complete vector velocity, obtainable from equations 2 and 7 is

$$v=\hat{x}(\Omega=\omega_L)y+\hat{y}(\Omega+\omega_L)x+\hat{z}s(z) \qquad (22)$$

From this expression it is apparent that the projections of the trajectories on the x-y plane are conic sections whose characteristics depend on the relative magnitudes of Ω and ω_L. When $\omega_L<\Omega$, the projected trajectories are hyperbolas with the common asymptotes

$$y=\pm\left(\frac{\Omega+\omega_L}{\Omega-\omega_L}\right)x \qquad (23)$$

and when $\omega_L>\Omega$, they are a one parameter family of congruent ellipses, with eccentricity

$$e^2=\left(\frac{2\Omega}{\omega_L-\Omega}\right) \qquad (24)$$

Of particular interest are the limiting cases $\Omega=0$ and $\Omega=\omega_L$. In the former, the projections are circles, and in the latter, they are straight lines.

The particular case in which s is the constant, $(eJ_0/2\epsilon_0 m\omega^2)$, and the trajectories project as ellipses constitutes a generalized Brillouin beam. As Ω takes its limiting values, the Brillouin sheet beam or cylindrical beam result. For the general case, the current density and charge density take values intermediate between the corresponding values for the limiting cases.

Potential Outside a Finite Beam

In this discussion it is assumed that $\omega_L>\Omega$, and that the beam is inside an elliptic cylindrical surface whose major and minor axes are prescribed. In addition, it is possible to specify the normalizing velocity of expression 20 and the current density. The potential inside the beam is then conveniently written in the normalized form

$$\phi=A(u^2+v^2/\beta^2)+w^2 \qquad (25)$$

where

$\beta=(b/a)^2$
$u=(x/a)$
$v=(y/a)$
$A=(eJ_0a^2/\epsilon_0 mv_n^3)(\beta^2/\beta^2+1)$
$w=s/v_n$
$a=$ minor axis of ellipse

From equation 25 it appears that the traces of the equipotential surfaces on planes normal to the z-axis are congruent ellipses with eccentricity equal to the square of the eccentricity of the trajectory ellipses.

That the finite beam described can be maintained, the potential and its normal derivative must be continuous across the beam edge, and outside the beam the potential must satisfy Laplace's equation. The analytic statement of this boundary value problem is most simply made in terms of the co-ordinates of the elliptic cylinder so that the beam boundary is described by a co-ordinate surface. The co-ordinate transformations are the following:

$$x=a\sqrt{\beta-1}\sinh\xi\sin\eta$$

$$y=a\sqrt{\beta-1}\cosh\xi\cos\eta \qquad (26)$$

$$z=z$$

The domain exterior to the beam is $\xi_0\xi\leqq<\infty$ and $0\leqq\eta\leqq 2\pi$. The beam boundary is described by the co-ordinate surface $\xi={}_0\xi$, where

$$\xi=\cosh^{-1}\sqrt{\beta/(\beta-1)} \qquad (27)$$

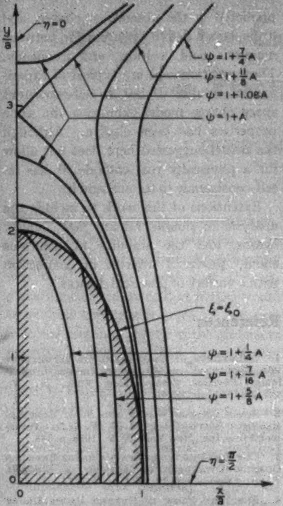


Fig. 2. Equipotentials for elliptical Brillouin beam with $(b/a)^2 = 4$

In terms of the new co-ordinates, expression 25 reads

$$\phi = A(\beta-1)\,[\sinh^2\xi\,\sin^2\eta + (1/\beta^2)\times \cosh^2\xi\,\cos^2\eta] + w^2 \qquad (28)$$

From equation 28 and its derivative with respect to ξ, the potential and its normal derivative on the beam boundary are obtained in the forms

$$\phi(\xi_0,\eta,z) = \frac{A}{2}\left(\frac{\beta+1}{\beta}\right) + w^2 - \frac{A}{2}\left(\frac{\beta-1}{\beta}\right)\cos 2\eta \qquad (29)$$

and

$$\left.\frac{\partial\phi}{\partial n}\right|_{\xi=\xi_0} = \frac{A\sqrt{\beta}(\beta^2+1)}{\beta^2}\left[1 - \left(\frac{\beta^2-1}{\beta^2+1}\right)\cos 2\eta\right] \qquad (30)$$

It is now required to find a solution in the exterior domain of Laplace's equation in the form

$$\frac{2}{a^2(\beta-1)(\cosh 2\xi - \cos 2\eta)}\left(\frac{\partial^2\psi}{\partial\xi^2} + \frac{\partial^2\psi}{\partial\eta^2}\right) + \frac{\partial^2\psi}{\partial z^2} = 0 \qquad (31)$$

whose value and normal derivative converge to expressions 29 and 30 respectively on the beam boundary.

The Cauchy problem as stated possesses a unique solution, but little is gained by writing it down. The solution is obtained by virtue of the separability of equation 31, and it is expressible in terms of a double infinite sum involving Mathieu functions of the $ce_{2m}(\eta)$ type, the associate Mathieu functions of the first and second kind, and cosine functions. The constants are related to those appearing in equations 29 and 30 and to the Fourier coefficients of the function $w^2(z)$, but the task of determining these constants is hopelessly complicated.

More enlightening than the general solution are the solutions for some particular cases. These suggest properties of the general solution, and are of some practical interest in themselves.

Potential of Generalized Brillouin Beam

If the axial velocity of the beam is constant, the normalized potential need have no z dependence, and equation 31 reduces to the 2-dimensional Laplace equation. In accord with equations 29 and 30, its solution is

$$\psi = 1 + \frac{A}{2}\left(\frac{\beta+1}{\beta}\right) + A\left(\frac{\beta^2+1}{\beta\sqrt{\beta}}\right)(\xi-\xi_0) - \frac{A}{2}\left(\frac{\beta-1}{\beta}\right)\left[(\cosh 2(\xi-\xi_0) + \left(\frac{\beta+1}{\sqrt{\beta}}\right)\sinh(\xi-\xi_0)\right]\cos 2\eta \qquad (32)$$

Several equipotentials for a beam of eccentricity $\sqrt{3}/2$, $(\beta=4)$, are shown in Fig. 2. Near the beam, the equipotentials are very similar to ellipses with the same eccentricity as the elliptical equipotentials in the interior of the beam. At greater distances from the beam, the equipotentials are no longer closed curves. All the curves of this class must ultimately become asymptotic to the lines $y = \pm x$. The potential at the saddle point shown is only a small amount greater than the potential on the beam edge where it cuts the x-axis.

From the potential plots it is apparent that a variety of electrode shapes and potentials for the maintenance of such a beam are possible, and the electrolytic tank should yield design shapes which can be readily machined. For the very long ellipse, which approaches the Brillouin sheet beam, it appears that electrodes at more than one potential are a necessity if the boundary conditions at the end of the beam are to be satisfied.

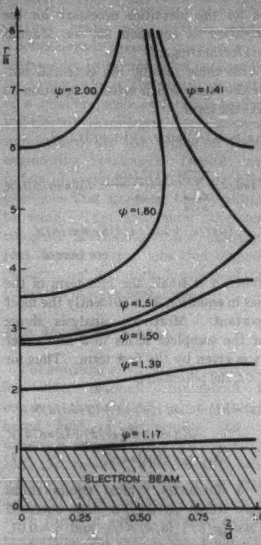


Fig. 3. Equipotentials for circular periodic beam with $(a/d) = 1/4$, $\gamma = 0.01$, and $A = (\pi/8)^2$

Potential of Circular Beam with Varying Axial Velocity

If the cross section of the beam is circular, considerable simplification of the general problem is achieved. (This case is discussed from a different point of view by Dow.[8]) It is appropriate to express the potential in terms of circular cylindrical co-ordinates so that by separation of variables a solution is obtained in the form of a sum over Bessel and trigonometric functions. The solution must satisfy the boundary conditions

$$\psi(a,z) = A + w^2(z) \qquad (33)$$

$$\left.\frac{\partial\psi}{\partial r}\right|_{r=a} = \frac{2A}{a}$$

Before writing the expression for the potential, it is necessary to determine the Fourier coefficients of the function $w^2(z)$. This may be done with an elegant result for the curtate cycloid of Fig. 1, although the problem is of considerable difficulty for the prolate case. With the assumption that $\gamma < 1$ and that

$$w^2 = \frac{1}{2}a_0 + \sum_{n=0}^{\infty} a_n \cos\frac{n\pi z}{d} \qquad (34)$$

it follows that

$$a_n = \frac{2}{\pi}\int_0^{\pi}(1+\gamma\cos\theta)^3\cos(n\theta + \gamma n\sin\theta)\,d\theta$$

$$a_0 = (2+3\gamma^2) \qquad (35)$$

$$a_n = -(-1)^n\frac{4}{n^3}J_n(n\gamma)$$

For the notation on Bessel functions

and for the identities necessary for the derivation of equation 35, see Magnus and Oberhettinger.[9]

With these results, the potential outside the beam which satisfies equation 33 takes the form

$$\psi(r,z)=A(1+2\log r/a)+1+(3/2)\gamma^2$$
$$-(4\pi a/d)\sum_{n=0}^{\infty}\frac{(-1)^n J_n(n\gamma)}{n}K_1(n\pi a/d)I_0\times$$
$$(n\pi r/d)+I_1(n\pi a/d)K_0(n\pi r/d)\times$$
$$\cos(n\pi z/d) \quad (36)$$

When γ is small, the first term of the series in equation 36 is evidently the most important. Moreover, analysis shows that the complete series to a first order in γ is given by the first term. Thus for $\gamma<<1$, the potential is

$$\psi(r,z)=A(1+2\log r/a)+1+(2\gamma\pi a/d)\times$$
$$[K_1(\pi a/d)I_0(\pi r/d)+I_1(\pi a/d)\times$$
$$K_0(\pi r/d)]\cos(\pi_n/d) \quad (37)$$

Fig. 3 illustrates typical equipotential surface traces on the r-z plane obtained for $A=(\pi/8)^2$, $(a/d)={}^1/_4$, and $\gamma=0.01$. The small value of γ is chosen to show the existence of equipotentials bounded with respect to r. For a larger γ, the curves which are nearly semicircular would be closer to the beam. For larger potentials than those shown, the equipotential curves approach a rectangular shape. It appears that the design of electrodes to produce this type of beam should be a simple problem even if it is desired to incorporate such electrodes into a slow-wave structure.

Conclusions

The exact solutions presented which are characterized by $\gamma<1$ are descriptive of regions containing no cathode, for there is no point at which the velocity vanishes. Nevertheless, the launching of a beam with elliptical cross section and periodically varying axial velocity should be possible by employing the techniques developed for the production of the circular Brillouin beam. When $\gamma=1$, the region may contain an even number of points at which the velocities vanish, and these may be considered as real or virtual point cathodes. Unfortunately, the solutions yield no surface of vanishing velocity and hence no model of a physically realizable cathode.

Although the generalized Brillouin beam possesses no distinguishing practical advantages over the circular Brillouin beam, it appears that the problem of producing a sheet beam might be approached to advantage in terms of the elliptical beam with large eccentricity. Since the sheet beam in practice cannot be semi-infinite, the elliptical model is in this respect a more realistic one.

A theoretical question which Brillouin's paper leaves unanswered is whether or not the circular beam is a particular solution of the general problem whose properties are dependent on the highly symmetrical cross-sectional shape. A similar question has been raised concerning the TE_{01} mode in a circular waveguide, and in this case it is well known that a distortion of the boundary into an ellipse eliminates the mode altogether. In the case of the circular Brillouin beam, however, the analysis presented here shows that it belongs to a 1-parameter family of solutions and is not a singular case.

Another possible application of these results is to the Barkhausen-Kurz oscillator. In recent work on this complicated device, much effort has been spent in attempting to find a satisfactory space charge model of the beam. The solution presented above with $\gamma=1$ and $\beta=1$ has the possibility of being a good space charge model, because it predicts a reduction in the frequency below that predicted in the absence of charge, and it displays a definite relationship between magnetic field, voltage, and frequency. These relationships are typical of experimental results,[10,11] but a self-consistent space charge model exhibiting the same properties has been elusive. Although the model suggested here does not allow for a physically real cathode, it has its self-consistency to recommend it.

Extensions of this work to include the analysis of double stream flow, hollow beams, and the stability of solutions should prove of interest in connection with a variety of practical devices.

References

1. A THEOREM OF LARMOR AND ITS IMPORTANCE FOR ELECTRONS IN MAGNETIC FIELDS, L. Brillouin. *Physical Review*, New York, N. Y., vol. 67, 1945, pp. 260–66.

2. SPACE CHARGE WAVES AND SLOW ELECTROMAGNETIC WAVES (book), A. H. W. Beck. Pergamon Press, Inc., New York, N. Y., 1958, p. 348.

3. CURRENT LIMITED BY SPACE CHARGE BETWEEN COAXIAL CYLINDERS, I. Langmiur, K. Blodgett. *Physical Review*, vol. 22, 1923, pp. 347–57.

4. ELECTRON FLOW IN CURVED PATHS UNDER SPACE-CHARGE CONDITIONS, B. Meltzer. *Proceedings*, Physical Society, London, England, section B, vol. 62, 1949, pp. 431–37, 813–17.

5. CONGRUENT SPACE CHARGE FLOW, G. B. Walker. *Ibid.*, vol. 63, 1950, pp. 1017–27.

6. CURVILINEAR SPACE-CHARGE FLOW WITH APPLICATIONS TO ELECTRON GUNS, P. T. Kirstein. *Report no. 440*, Stanford University Press, Stanford, Calif., Jan. 1958.

7. USE OF THE ACTION FUNCTION TO OBTAIN THE GENERAL DIFFERENTIAL EQUATIONS OF SPACE CHARGE FLOW IN MORE THAN ONE DIMENSION, K. Sangenberg. *Journal*, Franklin Institute, Philadelphia, Pa., vol. 232, 1941, pp. 365–71.

8. ADVANCES IN ELECTRONICS AND ELECTRON PHYSICS (book), W. G. Dow. Academic Press, Inc., New York, N. Y., 1958.

9. FORMULAS AND THEOREMS FOR THE SPECIAL FUNCTIONS OF MATHEMATICAL PHYSICS (book). W. Magnus, F. Oberhettinger. Chelsea Publishing Company, New York, N. Y., 1949.

10. A MAGNETICALLY FOCUSSED RETARDING FIELD OSCILLATOR OF THE BARKHAUSEN TYPE, M. Uenohara. *Technical Report no. 1*, Ohio State Research Foundation, contract no. AF18 (600)–982, Sept. 1954.

11. A THEORETICAL AND EXPERIMENTAL INVESTIGATION OF A TYPE OF BARKHAUSEN-KURZ OSCILLATOR, W. Simon. *Ph.D. Thesis*, Harvard University, Cambridge, Mass., Apr. 1958.

Education for Expanding Horizons in Electric Power

GORDON S. BROWN
FELLOW AIEE

TO RECALL the undergraduate education of the writer's generation is to see evidence that engineering education was based on the belief that the engineering curriculum of that period would serve him throughout most of his professional lifetime. Many of the professors teaching at that time had a firm idea of what their students needed to know in order to engage in careers in the professional world. This seemed appropriate, for these professors had in no small degree conceived most of the world of electrical technology of about that era. It looked sophisticated and enduring. They accepted the principle of a curriculum largely based on contemporary practice, and assumed that the engineering profession would merely build the mechanisms of society, often to a conventional pattern, and keep them running. This doctrine stemmed from the assumption that technology would not change appreciably during an engineer's professional career.

Today the problem for the engineering degree-granting college is different. The last two decades or more have witnessed great productivity in science with the uncovering of tremendous amounts of wholly new knowledge about the physical world. In our turbulent technological scene, engineering proliferates with accelerating tempo in many fields, each with a broad range of work function, each undergoing rapid technological change. The changing technology presents educators with an ever-widening spectrum of problems. With each successive college commencement, it becomes increasingly difficult as well as increasingly unwise to equip graduates with specialized know-how skills related to contemporary engineering practice. Not only will tomorrow's engineer need to know more than his counterpart of a few decades ago, but he needs to be prepared to learn more during his professional lifetime. And what he must learn will often be quite different from what he was taught in college.

Within the young and rapidly changing field of electrical engineering is a wide variety of work function. The majority of graduates work for industry and not for a client. Matters of public safety, of equipment cost, etc., are important, but the yardstick by which their importance is measured is a function of determining who is the customer. Where the customer is a branch of the military, as is often the case, the yardstick for performance and cost of devices and systems is often calibrated in terms of national survival rather than by a 6% return on investment, and change is often encouraged for change's sake. Modern science exerts a dominant role in all engineering, and especially in defense engineering. Much engineering is done prior to a demonstration of practical feasibility.

The contrast in criteria that set engineering values in: 1. the design of a multimegawatt generating station for a public utility, and 2. a multikilowatt generating plant for a large vehicle to operate in outer space, puts the picture into focus. The first system may be typified as stationary, coal-, oil-, or gas-burning for a uniform environment, requiring 100% reliability, low capital cost, and low maintenance. Each new plant must be superior to its predecessor because of a number of circumstances, but it comes into existence largely through an orderly integration and some extension of the known art of the day. The second system, on the other hand, presumably must have negligible weight and size, must use electromagnetic radiation as its source of energy, withstand large accelerations, and operate for limited time in environments never yet encountered by man. Furthermore, its design must be compatible with the changing needs of many interlocking requirements set by specialists in nagivation, guidance, aerodynamics, etc.

The engineers who are to create this system encounter many problems for which there is little prior art to supply many of the answers. They must lean heavily on fundamentals of science and on their ability to apply them.

At present, there is the gradual substitution of fission fuel for fossil fuel, with its attendant highly technical problem of protection of personnel and equipment against radiation. The future shows promise of wholly new techniques for converting heat into electric energy. Thermoelectric power conversion by thermopiles using advanced semiconductor technology, or thermionic generators on a grandiose scale, or fuel cells as chemical-electric converters, or thermofusion energy, are live topics on today's engineering campuses. Digital computers of amazing capability are taking over jobs that engineers once performed by hand. They and many other technological developments are exerting a vast impact on the jobs that engineers perform, on the work function of vast categories of professional men.

The Academic Scene

The academic environments in engineering differ markedly throughout the country; they also differ even within the broad field of engineering. To cite merely a few differences, some institutions are oriented toward research, others toward applied engineering or manufacturing. The big differences in the work function among different fields of engineering and even within fields, exert considerable influence on educational policy.

As engineering as a vocation and as a profession has evolved, a system has been devised of 2- to 3-year technical institutes, 4- to 5-year baccalaureate-degree-granting engineering colleges, a system for on-the-job education of graduate engineers in industry, and an expanding program of graduate education in engineering colleges. The doctor's degree is becoming a prerequisite for leadership in many engineering areas.

Since it appears certain that the 4-year program of undergraduate education in engineering in most colleges and universities will be adhered to for quite a while, engineering curriculums are allocating less time to a treatment of contemporary engineering practice and more time to broad basic science and to the engineering of science.[1–3] As professors concern themselves with this limited task while continuing to point their interests toward engineering, they find the available time for undergraduate education much too short for the job that industry and society expect of them. On all sides one sees a reassessment of the content of curriculums and the methodology for giving an en-

---

Paper 60-17, recommended by the AIEE Education Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., January 31–February 4, 1960. Manuscript submitted July 6, 1959; made available for printing October 7, 1959.

GORDON S. BROWN is with the Massachusetts Institute of Technology, Cambridge, Mass.

during, purposeful education to future engineers.

The problem before engineering educators involves being sensitive to the expanding body of scientific knowledge and to the wide spectrum of work function in industry. Educators must continually update the scientific or technical content of their programs of instruction. Likewise, industrial leaders must keep abreast of the changes that occur in college environments as, year by year, the professors' interests change with the unfolding of new knowledge, and as curriculums, in the updating process, become more broadly based in the sciences. Updating can no longer be achieved merely by adding new material to an existing curriculum. The problem is not what to add but what to drop in order to make room for change. Often the procedure is misunderstood, as evidenced by the comment of a British educator who recently visited the writer's office. He had noted the preponderance of electronics, and was most disturbed that the teaching of electric power engineering had nearly disappeared from the American campus. He had not observed, however, that everything electrical is a manifestation of charge carriers, that a broadly based program called electronics can embrace power or, more properly, electric energy processing; and that electric power itself is emerging in new, more modern, forms.

It appears, therefore, that the solution at the undergraduate level is to bring about, continually, fresh course syntheses that have wider applicability within the criterion of knowledge that will be negotiable at par in many allied fields for a reasonably long time. Specialization is being moved to the graduate level. But the guideposts for doing this are varied and often appear contradictory.

Guideposts to an Engineering Curriculum

That engineering is the exploitation of science for the betterment of man is widely accepted. In other words, engineering is applying science or practicing the quantitative organization of natural phenomena. But in its creative aspects it comprises the ability to relate seemingly unrelated events of nature in quantitative ways in devising useful configurations of matter, elements, devices, or systems. Engineering must also quantitatively classify the recorded observations of science in ways that will provide a prediction of the quality and quantity of the outcome of designs not yet concluded,[4] in order that achievement may result from suggestions, that the possibility of new applications shall be enhanced and the validity of theories extended to include a continually widening scope. One may say that doing engineering is practicing the art of the organized forcing of technological change.

But not all persons whose careers are in engineering are endowed with identical talents. Because of the wide diversity among people, the wide diversity of activity within engineering, and the limited scope of most engineers' activities, it is helpful to classify broadly the work functions of engineers. There are about as many work functions as there are people but to point up some meaningful aspects of activity in relation to education, one may try to classify engineering practice in three categories, knowing well that the specific kinds of competence sought are a blend of an engineer's natural gifts, what he learns in college, and what he learns on the job or just by living longer and observing the world around him. Though oversimplified, the following is submitted to identify broadly three categories of work function.

1. IMAGINATIVE INSIGHT

In this category belong the persons who have talent for detecting deficiencies in the capabilities of existing engineering devices or systems, and who can conceive from abstractions or from the whole body of science, wholly new superior devices and systems. These can relate quite abstract and seemingly unrelated concepts in useful ways. They are composers rather than arrangers. Their work is predominantly intellectual and varied. These are the engineers who, having perceived, for example, the need for increased capability in computation, can conceive and point the way to the giant digital machines long before the first one is built; or who, from an assessment of the dwindling fossil fuel supply, perceive the possibility of, and point the way to, nuclear power or the direct conversion of cheap heat into usable electricity by thermoelectric generators, or who can synthesize from abstract science the techniques by which man may tame the energy of thermofusion. The power industry needs these men within its fold.

2. CREATIVE INVENTION

In category 2 are the persons who can devise and build systems of engineering works by an orderly, creative, ingenious arrangement of the knowledge of the day. These men are more creative arrangers or inventors than composers. They work at a less abstract level than do the men in the first category; they are sensitive in a quantitative way to engineering feasibility, and to economics. They need a high degree of competence in the areas of mathematics, physics, and related sciences, for they must understand both how and why machines work, and perform complicated analyses. They design and build complicated engineering works and in a great many instances they are doing something new. Within their area of work, they may perform at the same level of creative achievement, but their achievement will be of a different character from that of the engineers in category 1. Without these persons the contributions of the engineers of category 1 would be slow in coming into physical actuality.

3. PRACTICAL SKILLS

Persons who are expert in assembling, operating, and maintaining complicated machines and engineering works from drawings, descriptions, and from experience are in this category. These are not second-rate citizens. In tomorrow's era of technology they will be needed in great numbers. They find it difficult to move ahead with the stream with less than 2 years of study beyond high school. The environment in which they learn, however, can appropriately be adjusted to teaching them the how in addition to the why of contemporary engineering practice, devices, and systems.

For the engineers of either category 1 or 2, the Bachelor's degree should be only a beginning. During their professional lifetime, these men will witness great change and will be required to learn a great deal of technology not even visualized during their college careers. Their careers will be influenced by the fact that never is a machine devised which, within a few years, is not called upon to perform at a higher level of capability than its creators built into it. Throughout the undergraduate program of these men, the emphasis should be on providing a sound training in basic science and in applying scientific method to the solution of tomorrow's problems rather than yesterday's, and on developing qualities of character and mind typical of the industrial leader. The curriculum should be pointed toward graduate work; for many, toward the doctorate. The environment should be charged with creativity. The knowledge acquired should be intellectually open-ended so that it will be negotiable at par in many fields for a long time to come. On the other hand, it may be appropriate to offer for the engineer in category 3, a Bachelor's degree program in a technical school and to regard it as terminal. His continued education will occur largely on the job.

The Field

The relative need for persons in these categories of work function varies in different fields and in different societies. In many societies the person in category 3 is the most valuable and most in demand. But regardless of where a person gets his education, there is no denying that he should be given the opportunity to develop to the highest intellectual level that his talents will allow. Unfortunately, our educational system does not find it easy to do this.

Some Specifics of Engineering Curriculums

Speaking specifically with respect to technical competence, and perhaps paradoxically, specialized professional competence in the newest fields is secure only when it is based on a sound fundamental training in the oldest of our disciplines, namely, physics, mathematics, and chemistry. In the newest fields, it should be emphasized again, within the 40-year professional lifetime that is ahead of today's college student, there will be many changes. Many new fields will emerge and demand their services.[6]

Hence, it is not surprising that in planning engineering curriculums, increased emphasis is being given to these old disciplines, and to building upon them new syntheses of a body of knowledge called engineering science. These engineering sciences couple the classic scientific subjects with new engineering frontiers. Education's new role is to illustrate the *engineering of science.* Indeed, it is this frontier coupling, or interaction in engineering, which must be made clear to students who want to become engineers and not scientists. The teaching of mechanics and dynamics in the broad sense; thermodynamics and statistical mechanics, mass momentum and heat transfer, electrical science of a kind that treats electric and magnetic field and circuit phenomena at an advanced level as related to energy and information processing; and microscopic properties of matter in relation to materials in engineering should be made relevant to tomorrow's engineering. No two colleges, nor for that matter any two departments of the same school, are approaching the future in an identical way. This seems proper in view of the wide differences among engineering fields and among academic environments. But the trend toward engineering science[3] in lieu of specialization for an undergraduate degree, has been noticeable long enough to indicate that for many institutions the development is beyond the point of no return.

As the increased relative importance of engineering science has stimulated college faculties to remake their course structures, there has been a widespread search for guideposts. Initially, the reassessment is limited to studies within departments; but pressure to do this on interdepartmental and interdisciplinary bases is increasing. Many educators question the wisdom of preserving an undergraduate program that is organized along conventional lines of electrical, mechanical, chemical, aeronautical, or metallurgical engineering, etc. The careers of many engineers prove to be no respecters of such categorization.[5] The existence of large digital computers as commonplace engineering tools has had a major influence on current engineering practice.[6] The boundaries between established departments on academic campuses are becoming more permeable, but professional society structures, the Engineers' Council for Professional Development accreditation practices,[7] licensing practices for engineers, etc., are so well entrenched and are such important parts of our heritage, that change is going to be evolutionary rather than revolutionary.

The criterion that "the knowledge should be negotiable at par in a wide range of allied fields for a significant period of time" should be a dominant factor in establishing guideposts. For an electrical engineer, another guidepost is found in the statement that everything that is characterized as electrical is a manifestation of the phenomena of charge carriers, their associated fields, and how they interact with materials.[8] To inject purpose into a program of study, it is helpful to say that electrical engineers exploit this nucleus or model of natural phenomena to perform useful functions in everyday living, either to process energy for energy's sake, or to process information so that humans can communicate. The processing of energy relates to man's desire to substitute machine for human muscle by providing the giant motors, machines, and power systems that control wheels of industry and also provide light and heat. The processing of information embraces techniques for establishing communications between man and man, between man and machine, and between machine and machine. Radar, radio, and wire telephony, measurement, electronics, computation, and control are the fields where information processing is applied.

First-rate competence in engineering practice is made more secure when an engineer is quite knowledgable in both broad areas. The above concept of electrical technology is being implemented by curriculums and teaching programs in American colleges in several forms. Behind the names of courses, one can observe emphasis on such open-ended topics and concepts as Electric Circuit Theory; Electronic Devices and Circuits; Information-Handling Systems; Electromagnetic Theory; Electrical Energy Conversion, Transmission and Radiation; and Modern Physics and Materials.

In many colleges, the sophomore and junior year work is substantially in science and electrical science. In the final year, students take various electives that couple the earlier study with areas of application. Here, hopefully, they come to grips with some of the new engineering frontiers.

Speaking strictly of the writer's own institution, the Massachusetts Institute of Technology has abandoned an option structure that was attempting to provide specialization by field, such as power, communications, computation, radar, etc. and is veering toward an option structure that shows signs of compatibility with categories 1 and 2 of work function, for they exist throughout all fields. For about 25 years, the Cooperative Course in Electrical Engineering, Course VI-A has been offered as one option. Recently, an Electrical Science and Engineering option, Course VI-B was initiated. The latter option is intended to provide the maximum opportunity for early development of the person whose eventual academic level is the doctorate and whose eventual work function would fit category 1. The program in Courses VI and VI-A fit students who may be oriented toward either categories 1 or 2.

In the evolution of this program, new areas of instruction have come into being and new laboratory facilities have evolved.

A graduate of the Department a decade ago will recognize in today's program little of what he studied, but it is hoped he may now see presented what he wishes he had studied. This is especially true of the graduate who has made his career in the electric power field. Alumni and educators from industry who know of the program in detail have given it a vote of confidence.

The Coupling of Industry and Education

As technology has advanced during the last decade, the newer fields have often appeared more glamorous than the older ones. For perfectly valid reasons, the research interests of engineering professors are in the newer fields. In fact, many of these fields were spawned by these research interests. As a consequence, the communications links between the older and well established fields and the new challenging activities on college campuses often become strained. In many instances, they have become disconnected.

This situation is not good. The field of power generation in the public utility area, while long established, continues to enlarge. It needs to recruit able young men to take care of growth, to stimulate its technological advance, and to replace the engineers who, after long years of productive service, have reached retirement age. It needs its proportionate share of first-rate minds in both categories 1 and 2.

The writer believes that ways must be found to restore this communication link, and that restoration will best be accomplished by activating open-ended on-campus research in energy-processing and related areas, rather than merely in electric power, in ways that capture the interest of both engineers in industry and engineering college professors.

It does not appear possible to establish research on engineering campuses around topics that relate closely to contemporary industrial practice. Faculty research is of questionable academic caliber unless it is basic, open-ended, and has with respect to its effect on industry, a lead time of several years. Some new horizons need to be explored, and this may be the time to do it.

For example, throughout all industry a major deterrent to advancement is the limited capability of existing materials. Basic research in the area of molecular engineering to achieve new materials and improvements in existing materials would be most fruitful.

Within the broad field of the direct conversion of heat into electric energy, there are many problems whose solution would yield useful engineering developments within 5 to 10 years. Specifically, the exploitation of solid-state technology, plasma dynamics, and magnetohydrodynamics to provide new machines to furnish electric power are good examples. Research in these areas should be flexible and imaginative, for no one living can visualize the technology of 10 years hence. The record clearly shows that the future will be different, that it will surprise us, and that the public utility industry will need what ever the future may bring.

The advent of the large digital computer[6] opens a wide vista for gainful research in the power area. Automatic monitoring of load generation and distribution presents a challenge. Contemporary digital computers possess a capability that so vastly exceeds the capability of network analyzers of merely a decade ago, that the methods for representing complete electrical generating systems on computers for stability and power-flow problems merit reassessment. It is no longer necessary to use oversimplified mathematical models of boilers, turbines, speed governers, generators, and voltage regulators.

As technologies for the direct conversion of heat into d-c energy reach the engineering phase, the following needs are apparent: 1. The invention and engineering of new machines to utilize energy in the form of kiloamperes at a few volts, in contrast with a few amperes at high voltage as used today. 2. Ingenious ways to convert large blocks of energy from direct current to alternating current, perhaps by devising ways to exploit modern silicon diodes and modern switching transistors. 3. Better and cheaper ways to convert from alternating energy into direct energy, and vice versa.

The long-term values of on-campus research in which a student participates need wider recognition. This does not imply that a student should be trained only for a job in research and development but, rather, to enhance the probability of developing his talents of creativity and versatility or originality. Even if he is not destined for a career in research and development, this may be a wise choice, for, increasingly, management leaders are being selected from personnel within this category.

The electric utility industry has yet to find a way to allocate funds to college professors for research of a professor's own choosing that is compatible with an academic career. The older procedure of contracting with colleges for programmatic work that is more engineering than research, and monitoring it from the central office, does not meet with the enthusiasm that is necessary. Young, imaginative students, enthusiastic about a career in the public utility field, do not favor the prospect of monitoring engineering which is practiced elsewhere. The result is that there is a poor impedance match between the industry and the campus that often results in many highly desirable students looking elsewhere for employment and hence overlooking a wonderful opportunity for an exciting career.

Conclusions

The persistent substitution of machine for human muscle forecasts continuing programs to construct electric generating and distribution plants to keep ahead of demand. With increased load comes a demand for increased size of plant with increased performance. Greater numbers of engineers are required to build bigger and better plants on a shortened time scale and to incorporate new technological developments. Concurrent with these trends is an exfoliation of engineering into many new highly technical and challenging fields that compete for engineers.

These situations tax the technical and administrative skills of the practicing engineer, and of the professor who is to educate tomorrow's engineer. Since engineering colleges are restricted, in the main, to the 4-year undergraduate program, they are adjusting to the challenge by presenting students a more integrated body of fundamental knowledge, at the sacrifice of the specialized knowledge characteristic of the past. Many graduates of a modern 4-year curriculum are capable of a career in any one of several industries. When an employer first talks with a new 4-year graduate, he is less likely to find a specialist ready for his industry than did his predecessor a decade or so ago; this is especially true of employers in the electric power industry.

Industry needs to keep abreast of the changing technical content of modern engineering curriculums, and of the changing educational environment, in order to communicate fluently with tomorrow's graduate. It needs to appreciate the vital role of on-campus research and graduate study in keeping our educational institutions as places of learning, rather than merely places for teaching. The importance of recruiting top-flight young minds into all segments of industry on a balanced basis needs attention. The importance of aiding practicing engineers to obtain refresher instruction in an organized manner every 5 years or so is increasing. The key to the much needed communication between the leaders in the electric power industry and the professors on engineering college campuses is to initiate more open-ended programs of on-campus research under direct industry sponsorship.

Vol.
Pt. 1
196

References

1. Evaluation of Engineering Education, Committee Report. *Journal*, American Society for Engineering Education, Urbana, Ill., vol. 46, no. 1, Sept. 1955.

2. Evaluation of Engineering Education, Special Committee Report. *Special Report*, American Society for Engineering Education, June 1958.

3. Recent Advances in the Engineering Sciences (book), George A. Hawkins, editor. McGraw-Hill Book Company, Inc., New York, N. Y., 1959.

4. Science and Engineering Education, R. B. Adler. *Journal*, American Society for Engineering Education, vol. 47, no. 2, Oct. 1956, pp. 121–28.

5. Demands of an Explosive Era, C. H. Linder. *Electrical Engineering*, vol. 77, no. 8, Aug. 1958, pp. 674–80.

6. Digital Computers for Distribution Systems, D. N. Reps. *Ibid.*, pp. 708–13.

7. ECPD Accreditation of Engineering Curricula, C. C. Chambers, G. F. Branigan. *Ibid.*, pp. 686–88.

8. Educating Electrical Engineers to Exploit Science, G. S. Brown. *Ibid.*, vol. 74, no. 2, Feb. 1955, pp. 110–15.

9. Education for Electrical Engineering, Lan Jen Chu. *Transactions*, Professional Group on Microwave Theory and Technique, Institute of Radio Engineers, New York, N. Y., July 1958, pp. 249–50.

◆

Discussion

V. F. Estcourt (Pacific Gas and Electric Company, San Francisco, Calif.): We note with interest that this more recent presentation of the paper (first presented in 1958) contains a number of significant revisions. The author is to be congratulated for this thoughtful review of a number of important educational fundamentals of methods for bringing about a closer relationship between industry and education.

We agree that the communication links between the older and well-established fields and the challenging activities on college campuses often become strained, and in many instances disconnected. In recommending ways for the restoration of this communication link, emphasis is placed upon the activation of "...open-ended on-campus research in energy-processing and related areas, rather than merely in electric power, in ways that can capture the interest of both engineers in industry and engineering college professors," and the author concludes that faculty research is of questionable academic caliber unless it is basic. Certainly we must agree that faculty research around topics that are too closely related to contemporary industrial practice is not only unattractive to teachers of engineering but, because such topics normally lie outside the field of engineering teaching, activity in this area would be largely unproductive of constructive results.

Throughout the paper, the need for basic research is emphasized as the key to communication between leaders in the electric power industry and the professors of engineering college campuses. This is good to the extent that it will bring significant advances in the art, or justifiable economic returns in the short or the long range. But the author stops there and fails to mention other desirable links. To this extent it becomes difficult to see the major problem in its proper perspective.

It is natural that research efforts by power engineers are predominantly in the field of applied rather than more fundamental research. This does not mean that the power industry lacks interest in this field, but merely that educational institutions, special research organizations, and large manufacturers usually provide a more suitable climate for the successful conduct of this type of activity. On the other hand, it should be pointed out that a most intimate relationship exists between the engineers of a good many utilities and the major power equipment manufacturers, and the two groups are constantly teaming up in co-operative research efforts.

The engineering scene in the power industry has been changing rapidly, but engineering teachers in some instances appear still to be thinking in terms of situations which existed 10 or 15 years ago. A great amount of applied research and development is continually going on in, or in relation to, existing power plants, as well as in connection with radically new designs, and the outlook is for further spectacular advances in the art, partly as a result of such activities. A large modern thermal power plant (conventional or nuclear) is in a certain sense one of the best equipped and most challenging research laboratories available for advanced engineering study. We suspect that many teachers of engineering would be surprised if they realized the full extent and significance of high-level research which goes on in connection with some of the large modern power plants. Utilities have perhaps been lax in bringing their story to engineering faculties, and a recent paper by this discusser represents an effort to establish a closer understanding between engineering colleges and the power industry.[1]

In the section "Guideposts to an Engineering Curriculum," in category 1, the author states that the kind of individual required is one "who can conceive from abstractions or from the whole body of science, wholly new superior devices and systems." This recognizes a fundamental objective in certain fields of engineering science, and is especially applicable where it involves the transition from one technological plateau to another. On the other hand, to certain types of intellects, each such plateau appears to a large extent as a distinct and static situation in the sense that it cannot be extended or improved in any significant amount unless it is completely displaced by other "wholly new and superior devices or systems." What must be understood is that, in many instances, each plateau of a major technological advance contains within it much of a dynamic nature which makes possible further expansion or development of radically new concepts leading to a significant technological advance, or sometimes a series of less pretentious advances which, in the aggregate, represent a major transition to a wholly superior device or system. Therefore, we must also recognize those individuals who have a feeling for the dynamic possibilities inherent in a given device or system, and who have the intellectual qualifications to develop radically new applications rather than wholly abandon one device or system for an entirely different one. Frequently these two approaches require substantially different types of intellect and motivation. Broadly speaking, this distinction emphasizes one of the differences between a scientist and an engineer.

The significance of this point of view is sometimes overlooked. Existing devices or systems are too often regarded by the abstract thinker as inherently static in the sense that all of the engineering has already been completed and wrapped up in a package which provides no room for further creative and significant technological contributions. It would appear that the author's thinking is in this channel when he states, in the section "The Coupling of Industry and Education," that young imaginative students, "enthusiastic about a career in the public utility field, do not favor the prospect of monitoring engineering which is practiced elsewhere." Doubtless, such situations may exist in varying degrees both in the utility field and in other heavy industries. Since jobs requiring the highest level of academic achievement may be limited in these industries to approximately 25% to 40% of the engineering payroll, mistakes are sometimes made in the allocation of personnel by not recognizing the need for a proper balance of skills. Academic quality and over-all quality are not necessarily synonymous, and this must be taken into consideration in the assignment of engineering personnel.

There is another side to the problem. It is something of an anomaly that, as engineering curriculums become more concentrated in the study of the fundamental sciences, the end result in industry is a very large increase in the number of specialists, which appears to be out of proportion to the inherent need with respect to problems definitely requiring such specialization in depth. Thus, although it might reasonably be assumed that the broader knowledge of fundamentals acquired by today's engineering students would develop the ability to encompass larger areas of engineering, the reverse is frequently the case. This is partly because of overemphasis upon rigorous logical processes of abstract thought and mathematical formalism which tend to place the world of reality out of focus. As Prof. Keenan has aptly observed in an unpublished paper, "Many young people who will ultimately be competent, even brilliant, members of the profession would find that such a program would inhibit rather than stimulate their creative talents." It would appear that educators in the field of engineering have an important obligation to guard against the danger of leading engineering students into areas of intellectual irresponsibility with too little regard for worth-while objectives in the field of application.

We can certainly agree with the author's timely admonition that "Industry needs to keep abreast of the changing technical content of modern engineering curriculums, and of the changing educational environment, in order to communicate fluently with tomorrow's graduate." We are aware that efforts by industry in this direction have sometimes left much to be desired. However, it is also reasonable to expect that engineering professors will continue to maintain adequate communication with industry in order to retain close familiarity with its technological advances and progressively unfolding objectives. Thus the challenging outlets for brilliant minds may be better understood by educational institutions. A number of engineering colleges have been eminently successful in maintaining realistic contacts with what the better power companies are doing in application and development, but by no means has this always been the case.

Brilliant scientists are understandably in demand for much needed basic research which frequently demands unusual ability in the area of creative abstraction. However, there appears to be too much of a tendency to think of brilliant engineers as scientists and the less brilliant ones as engineers. As I have stated, "We must get away from the idea, expressed or implied, that an engineer is essentially a scientist with some practical experience in the design of hardware, as a result of which he has acquired judgment in the execution of his assignments. This is a very dull and unrealistic concept of an engineer, but unfortunately it represents a viewpoint which is not uncommon in the field of education."[2]

In conclusion, we endorse wholeheartedly the suggestion that basic research can be made a natural and worth-while communication link between industry and engineering colleges. As the author observes, "This does not imply that a student is being trained only for a job in research and development." On the other hand, if no other links with industry are created in the mind of the student, it suggests a lack of something of profound importance either in the engineering curriculum or in the point of view presented to the student.

It is our hope that this discussion of the author's fine presentation, offered from the viewpoint of an engineer and technical manager in the power field, will be accepted as a contribution to his expressed objective of establishing a stronger "coupling of industry and education."

REFERENCES

1. AN ENGINEERING CAREER IN THE POWER FIELD, V. F. Estcourt. *Mechanical Engineering*, American Society of Mechanical Engineers, New York, N. Y., vol. 58, no. 6, Nov. 1958, p. 92.

2. See reference 1, p. 93.

Gordon S. Brown: I am indebted to Mr. Estcourt for his provocative comments on my paper, and agree with many of them. But I believe that there is still a breakdown in communications and that the discussion has not provided a solution.

I am well aware that there is good team play between the public utilities and the major equipment manufacturers, but I am pleading for interplay between the utilities and the college campus. Many of my colleagues have close association with public utility problems but find it difficult to bring much back to the classroom or laboratory in a way that makes it a growing part of academic life.

My emphasis on basic research, with due recognition of the "humpty-dumpty" nature of the word basic, is to make it possible for the professor and his students to do work that conceivably could lead to self-generating ideas. Mr. Estcourt appears to be critical of my Category I engineer, but I believe firmly that unless professors and selected students can work at the kind of problem characterized by this category of work function, the self-generated idea is unlikely to arise. It is the open-ended, self-generated idea that is ideally suited to being followed up in a college environment.

I do not deplore design in the broad sense, but I know of no way to accomplish public-utility contemporary design under authentic circumstances in a college environment. On the other hand, I strongly support the introduction of students in mature ways into design problems. It so happens that in almost all research problems in engineering there are abundant design problems that students can tackle, and which reflect bona fide situations.

Perhaps the crux of the issue is found in my interpretation of the role of a university or technical college in today's society. I believe that not only should it teach well the knowledge of the day but also it should provide an environment where the ideas and issues of tomorrow are being uncovered. Here it must give a strong emphasis to learning rather than teaching. It is not enough for professors to be abreast of contemporary public utility practice and problems; they should be forging ahead into new problems. Likewise, it is not enough to equip students with the knowledge of the day; they need the basic knowledge from which will spring the developments of the future. They should be given the attitude of mind that will tend to bring these developments into reality. If they are going to be leaders they will be the instigators of change in the interests of improvement in technology and our society in general. The engineer is the person who has shaped our past and he will shape our future.

As I stated in the paper, electric energy is a vital element of our industrial system. Electric power engineers of the future will not be limited to hydro or fossil fuels as the basic source. New sources, and new forms for the processing of energy are clearly in the offing. Now is the time to be educating a future generation of public utility engineers in anticipation of these new forms, for, if this is done wisely and well, today's students, by their efforts as practicing engineers, will bring into reality what is now over the horizon. As they do engineering they will be participating in the organized forcing of technological change. Perhaps we have here the means for man to achieve the future that today he wishes were already a reality.

Development and Application of Time-Division Multiplexing Telemetering Equipment for Data Transmission on an Electric Power System

C. P. ALMON, JR.
FELLOW AIEE

JOHN DONELSON, JR.
ASSOCIATE MEMBER AIEE

Synopsis: In the fall of 1957 the Tennessee Valley Authority (TVA) began preliminary work toward increasing the number of telemeter readings by use of existing telemeter circuits with a minimum of modifications and additions. The additional channels were used to provide the power-system dispatchers with data that were normally received by telephone. The initial objective was to transmit the load readings of five hydroelectric plants and their total from the Nashville Load Dispatching Office to the Chattanooga Load Dispatching Office. High-speed data transmission was not a requirement in this case, since no automatic controls were involved and the readings were to be used for informative purposes. In view of the foregoing requirement, studies indicated that distant readings could be transmitted to a remote central location, from which time-division multiplexing telemetering equipment could be employed for data transmission, thus reducing the cost of communication channel facilities. Investigations indicated that no suitable equipment was available for this purpose that would work with existing facilities and that would provide the necessary functions. TVA devised experimental equipment to transmit six hydro plant readings over one impulse channel from the Nashville Load Dispatching Office to the Chattanooga Load Dispatching Office. Later the hydro plant readings were replaced by six substation load readings that are used in a computer for system economic loading purposes. This equipment has operated over a 1-tone channel of a microwave system about 130 miles in length for more than $1^1/_2$ years and the operation has been satisfactory.

THE OPERATING AREA of the TVA covers approximately 80,000 square miles, comprising portions of the states of Kentucky, Virginia, North Carolina, Georgia, Alabama, and Mississippi and most of the state of Tennessee. The rapid growth of power generation and the distribution of electicity through this area have led to a need for transmitting more quantities from distant areas to the central dispatching office in Chattanooga for system operation and economic loading purposes.

In TVA's automatic load-frequency control system numerous interchange readings with other companies are transmitted direct by high-speed telemetering over carrier channels. The interconnecting companies' megawatt load readings are totalized in Chattanooga and compared with a schedule; and the deviation from schedule is transmitted to the appropriate generating station, which automatically adjusts its output to compensate for the deviations from schedule. It is of great importance that time delays be reduced to a minimum in this control operation.

There are other readings in which high-speed response is not required, such as those needed for general dispatching information and economic loading. With a reduction in the speed of data transmission the telemetering channels could be reduced by applying the principles of time-division multiplexing. With this system, kilowatt load information gathered from various locations at a distant subcenter could be sequentially scanned and transmitted over one narrow-band telemeter channel.

It was found that suitable equipment that would work with existing equipment and that would transmit the desired quantities was not available. Since the output from the receivers at the Nashville Load Dispatching Office was in millivolts proportional to power, direct application of these quantities into an electronic scanning device did not appear to be feasible at that time. Several methods of transmitting the megawatt readings were considered; however, studies indicated that impulse duration transmission was best suited for the purpose, since an existing Leeds & Northrup 2.4-second Speedomax impulse telemeter transmitter could be incorporated easily into the transmission system.

After it was decided to use impulse lengths proportional to power readings, other considerations were: the methods of scanning the local inputs, the synchronizing circuit for the transmitter and receiver, a comparator circuit to convert impulses into an electrical quantity, and a means of synchronizing the comparator circuit with the memory circuits that hold the readings constant until the next comparison is made. Fig. 1 is a block diagram of the systems operation.

The Transmitter Unit

Six quantities were initially selected for retransmission, and the inputs for these quantities were derived from retransmitting slide wires that were already in use in a totalizing circuit. Two telephone-type rotary switches were chosen for the transmitter unit—one operates as a scanner and the other as a timer for synchronizing. A 22-point 5-bank switch

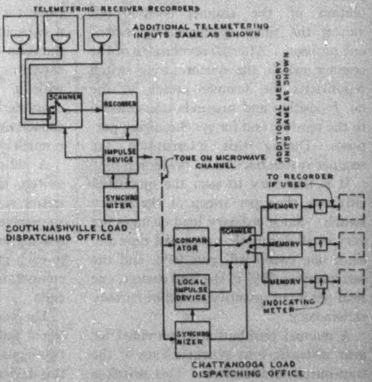

TELEMETERING RECEIVER RECORDERS
ADDITIONAL TELEMETERING INPUTS SAME AS SHOWN
SCANNER
RECORDER
IMPULSE DEVICE
SYNCHRONIZER
SOUTH NASHVILLE LOAD DISPATCHING OFFICE
TONE ON MICROWAVE CHANNEL
ADDITIONAL MEMORY UNITS SAME AS SHOWN
TO RECORDER IF USED
COMPARATOR
SCANNER
MEMORY
MEMORY
MEMORY
LOCAL IMPULSE DEVICE
INDICATING METER
SYNCHRONIZER
CHATTANOOGA LOAD DISPATCHING OFFICE

Fig. 1. Simplified block diagram of time-division multiplexing telemetering equipment

---

Paper 60-245, recommended by the AIEE Telemetering Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., January 31–February 5, 1960. Manuscript submitted November 4, 1959; made available for printing December 17, 1959.

C. P. ALMON, JR. and JOHN DONELSON, JR. are with the Tennessee Valley Authority, Chattanooga, Tenn.

The authors wish to acknowledge the assistance by Mr. Fairfax, Mr. Gostin, and Mr. Huff of the Tennessee Valley Authority.

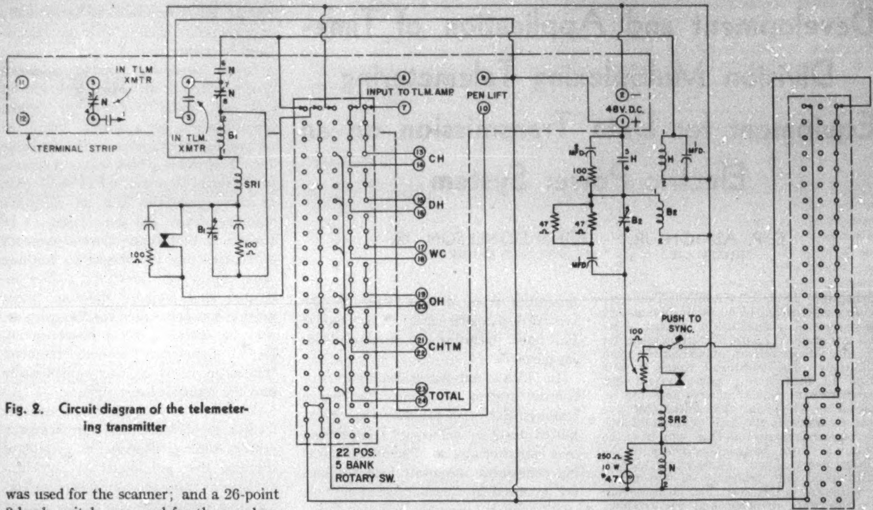


Fig. 2. Circuit diagram of the telemetering transmitter

was used for the scanner; and a 26-point 3-bank switch was used for the synchronizer timer. These switches were available and could be adapted to TVA's needs. Three positions are assigned for each reading. Postion *1* in each group is connected at the transmitting end only and is used to allow time for the transmitter recorder to achieve a balance. Postion *2* transmits 91% of the change since the last scanning. Position *3* transmits 99% of the change.

The start, or leading edge, of the time-duration impulse causes the scanner to move one position. The scanner switch also controls the pen lift of the transmitter recorder so that a dot is made for each quantity being recorded, operating very much in the same way as a multiple point temperature recorder. Each revolution of the scanner switch advances the timer synchronizing switch one position. After 20 revolutions of the scanner switch, the synchronizing switch repositions the scanner switch to the no. *1* position and suspends transmission to the receiver end for synchronizing purposes. During this interruption the scanner switch remains on position *1*.

It is necessary to scan the quantities only one time per sweep of the rotary switch. Since ordinary load changes are very small during that period of time and since one sweep will give 91% and the second sweep 99% of the load change, one scan per input quantity would be entirely satisfactory.

A manual pushbutton is provided for near instantaneous synchronizing of the transmitter and receiver. This switch is used for testing purposes and maintenance.

Fig. 2 shows the circuit diagram of the transmitter. The front and rear views of the transmitter are shown in Fig. 3.

The Receiver

The receiver is composed of a comparator circuit; a 22-position 5-bank telephone-type scanner switch; a synchronizing circuit with a local impulse device; memory circuits; indicating meters; and dropping resistors for use as millivolt potentials for recorders.

The comparator circuit is a resistor capacitor circuit in which the charge on a capacitor is proportional to the time duration or length of the impulse. The capacitor used is of a high-grade type with a very low dielectric-absorption characteristic. The RC (resistance-capacitance) time constant of the circuit is much longer than the longest impulse.

A 22-position telephone-type rotary switch, the same as used in the transmitter, scans the memory circuits. Three successive positions are assigned to each reading, with the second and third points in each group being used to connect the comparator circuit with the memory circuits.

The synchronizing circuit consists of a signal failure device that operates when two signal impulses are suspended; and this device synchronizes a cam impulse de-

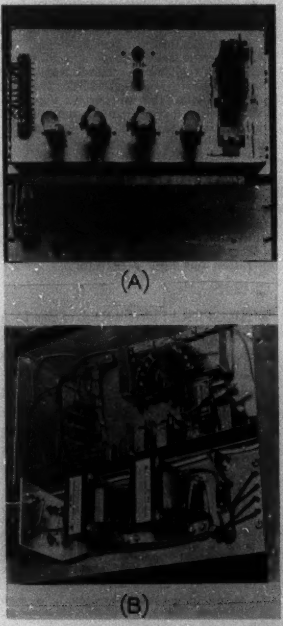


Fig. 3(A). Front and (B) rear view of transmitter unit

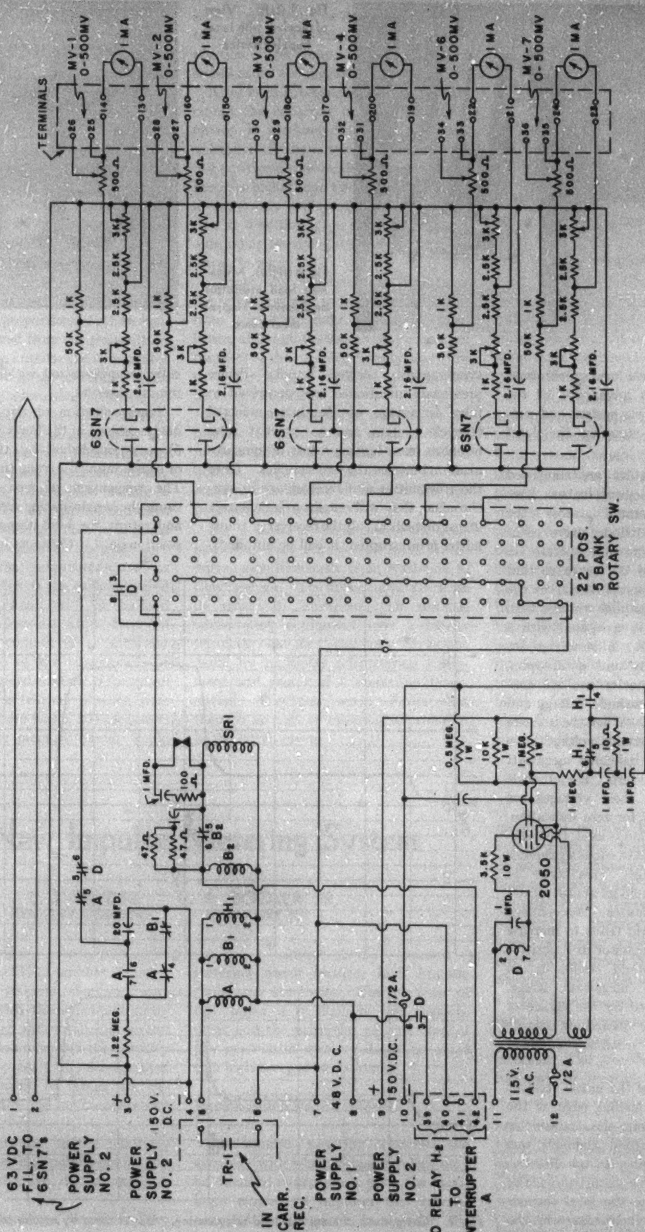

TERMINALS
MV-1 0-500MV
MV-2 0-500MV
MV-3 0-500MV
MV-4 0-500MV
MV-6 0-500MV
MV-7 0-500MV
1 MA
500 Ω
3K
2.5K
1K
50K
2.16 MFD.
6SN7
22 POS. 5 BANK ROTARY SW.
1 MFD.
100 Ω
47 Ω
B2
SR1
20 MFD.
A
B1
D
H1
1.22 MEG.
1/2 A.
0.5 MEG. 1W
10K 1W
1 MEG. 1W
10 Ω 1W
1 MEG.
1 MFD.
3.5K 10W
2050
115 V. A.C.
1/2 A
63 VDC FIL. TO 6SN7's
POWER SUPPLY NO. 2
POWER SUPPLY NO. 2 150 V. D.C.
TR-1
IN CARR. REC.
POWER SUPPLY NO. 1 48 V. D. C.
POWER SUPPLY NO. 2 150 V.D.C.
TO RELAY H2
TO INTERRUPTER A

Fig. 4. Circuit diagram of the telemetering receiver

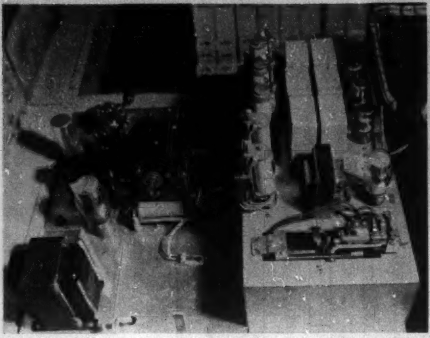


Fig. 5 (left). View of receiver with local impulse device

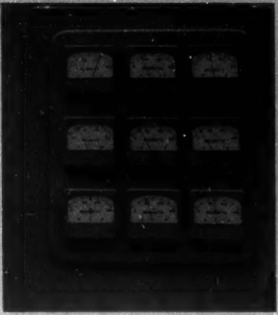


Fig. 6 (right). Megawatt load indicators on dispatcher's instrument board

vice that is driven by a local synchronous motor, which takes advantage of the interconnected 60-cycle system and supplies impulses for stepping the local scanner switch.

The memory circuits are composed of twin triode vacuum tubes having capacitors shunted across the input grids. These triodes have cathode dropping resistors and operate on the linear portion of the curve so that the plate current is proportional to the grid voltage. The cathode voltage with respect to negative is compared with a reference voltage. A milliammeter for indicating megawatts and a dropping resistor for potentiometer use are connected between the cathode voltage and the reference voltage. When the cathode voltage and the reference voltage have the same value, the indicated megawatt on the milliammeter is zero. If a left-scale instrument is used, these voltages are adjusted to be equal for zero indication. If zero-center operation is used, the cathode voltage is adjusted to equal a reference voltage at midrange. This type of memory circuit will hold an almost steady output for several minutes. A secondary emission from the grid tends to increase the charge on the capacitor in a positive direction, which would result in an increased indicated megawatt usage. This effect is nullified by the choice of capacitors that have dissipation factors about equal to the secondary emission of the grid circuits.

In the operation of the initial receiver circuit the start, or leading edge, of the received impulse from the transmitter over the communication channels was used to step the scanner switch directly. During communication disturbances false impulses would cause the local scanner switch to get out of synchronism with the transmitter scanner switch. The local synchronously driven impulse device previously mentioned was contrived to take advantage of the interconnected 60-cycle system and to prevent false impulses from causing a loss of synchronization between the two switches. Once the transmitter and receiver are in synchronism, they will remain so indefinitely unless disturbed. If the receiver or transmitter is interrupted, it will be automatically resynchronized within a preset 15-minute interval.

Synchronization of the local impulse device (shown in the lower portion of Fig. 5) is accomplished by the suspension of signal impulses from the transmitter. The suspension of two impulses will cause the synchronizing device to operate and return the local scanning switch to position no. *1*. Contacts in the synchro-

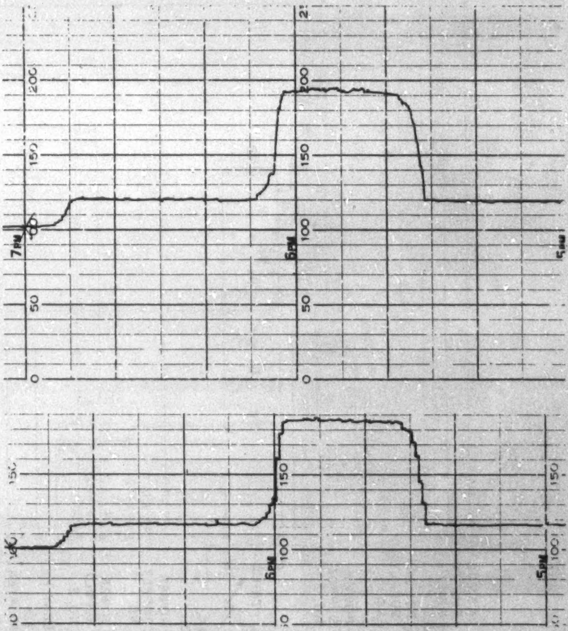


Fig. 7. Comparison of megawatt-load telemetering. Upper curve by regular telemetering and lower curve by time-division multiplexing telemetering

Vol.
Pt. 1
196

nizing device also operate a relay which locks in a notch of the motor-driven cam and stops this cam from rotating by means of a slipping clutch. When the leading edge of the first impulse is again received from the transmitter, the synchronizing device unlocks the cam, which starts rotating. The local impulse device steps the scanning switch to position no. *2*. At the same time the scanning switch in the transmitter is stepping to position no. *2*, after which both switches remain in synchronism.

During the operation of the receiver the capacitor in the comparator circuit is charged through a resistor, and the value of the charge is proportional to the length of impulse received from the transmitter. At the end of the impulse the comparator circuit is switched to the scanning switch. There are three successive positions on the switch assigned to each reading; and the first position in a group is not connected, since the transmitter recorder may not have reached a balance and any unbalance would cause a false reading in a memory circuit. The second and third points of the scanning switch in a group are connected successively to a memory circuit. Since the comparator capacitor is 10 times as large as the capacitor used in the memory circuit, 91% of the load change since the last scan will be recorded on the second contact of a particular group, and 99% of the change will be recorded on the third position of the group For instance, if a receiver position were reading 0 and there was a 100-megawatt load change, the memory circuit would read 91 megawatts on the second position of the group and 99 megawatts on the third position of the group. A successive scan would cause 100 megawatts to be read.

After each comparison with a memory circuit the capacitor in the comparator circuit is discharged. When the grid voltage in the memory circuit is increased in a positive direction, the cathode voltage becomes higher than the reference voltage as previously explained, and causes current to flow through a milliammeter and a dropping resistor, thus indicating an increase in load.

During any synchronizing operations the impulse circuit to the scanning switch from the comparator circuit is open; hence, although the switch may pass over contacts of the memory circuits, there are no spikes or false readings made in the memory circuits.

Fig. 4 shows the circuit diagram of the receiver. Fig. 5 shows a view of the receiver with local impulse device at the bottom. Fig. 6 shows a view of the megawatt indicating meters on the dispatching instrument board.

Fig. 7 shows a comparison of the same megawatt load curves that were transmitted simultaneously from Nashville to Chattanooga. The upper curve was made by means of 2.4-second-cycle impulse telemetering equipment over a carrier telemeter channel. The lower curve was made by multiplex telemetering equipment and was 1 of 6 quantities transmitted. The comparison shows that there is no loss of essential information in the latter curve.

Conclusions

1. The 2.4-second scanning period of the circuit is relatively slow, but it has been entirely satisfactory and is much faster than is required for the present purposes.

2. Each quantity is scanned at an interval of $45^{6}/_{10}$ seconds. By speeding up or reducing the balancing of time on the Leeds & Northrup potentiometer transmitter, the same circuit can transmit 19 quantities in the same length of time as is now being used to transmit 6.

3. The use of the system described has been very effective and reliable in transmitting megawatt readings of generating plant output and substation power usage from the remote center to the power system dispatchers' office in Chattanooga.

4. In addition to the present use the multiplexing system can be applied to transmit incremental-cost data on a far-flung system, where the distances from the dispatching center to the various plants are relatively large. With this system a single transmitter can transmit to a number of receivers located at the various plants over a single-frequency, narrow-band carrier channel; and only the receiver points applying to a particular station will appear at that station. For instance, the bus costs or lambda, divided by the penalty factor for each station, could be successively transmitted; and, with auction-type bidding for the system incremental load, each station could match its incremental cost with the incremental bus cost.

.79
1
50

A New Impulse Metering System

C. J. SNYDER
ASSOCIATE MEMBER AIEE

C. A. BOOKER, JR.
NONMEMBER AIEE

THE INCREASING number of power loads in the 100-mw (megawatt) and higher ranges with the attendant problems of metering and totalizing several supply sources has disclosed the inability of much of the present impulse-operated metering equipment to render the dependable, accurate, and maintenance-free service required.

This is in part due to the relatively low activity of such apparatus which has not justified extensive design improvement.

Recent developments, however, have made possible a new concept of impulse metering which features high impulse rates and accuracies. The purpose of this paper is to describe the development and the principles of operation of the new units now available to meet this high level of performance.

Performance Requirements

In impulse metering systems the accuracy with which demand values may be obtained is dependent upon the maximum number of impulses that can be accepted by a receiver or recorder, the number of circuits totalized, and the number of impulse totalizing steps involved which may harbor impulse storage errors. The obvious requirement for high accuracy is the capability for high impulse rates wherein each impulse represents a minimum kwhr (kilowatt-hour) or similar value so that storage errors in the meter contact devices and elsewhere are minimized.

The fundamental limit on impulse rates is the acceptance capacity of the terminal recorder in impulses per demand interval. The present trend is toward a full scale or printing capacity of 1,000

Paper 60-64, recommended by the AIEE Indicating and Integrating Instruments Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., January 31–February 5, 1960. Manuscript submitted November 2, 1959; made available for printing November 24, 1959.

C. J. SNYDER and C. A. BOOKER, JR. are with the Westinghouse Electric Corporation, Mr. Snyder at Raleigh, N. C., and Mr. Booker at Cheswick, Pa.

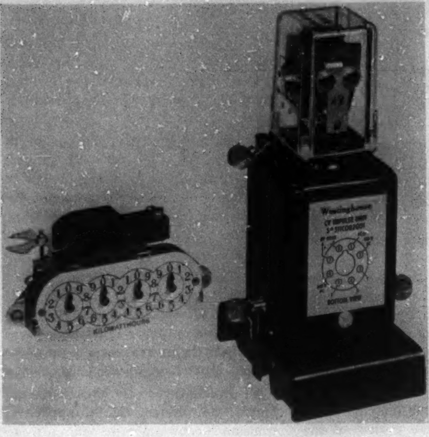


Fig. 1 (left). Oscillator impulse device

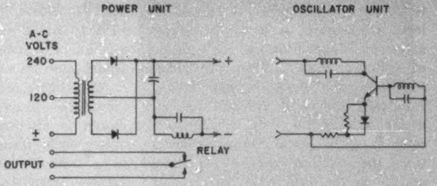


Fig. 2 (above). Impulse device schematic

impulses per interval. The ultimate accuracy of a recorder with 1,000 impulses per interval capacity is, of course, no better than ±0.1% at full scale. Such a recorder would normally be operated between 50% and 75% of full scale to provide adequate overload margin and the accuracy would therefore be more probably in the ±0.2% area. Since greater accuracy is certain to be desired in the future any new impulse system should readily handle about 5,000 impulses in a 15-minute demand period and have an inherent capability for even higher rates if required.

Another requirement is that of dependability, which is particularly important in the case of billing major revenue loads or interchange of power. This factor has in the past tended to restrict the use of electronic equipment and favor mechanical devices.

Any new system also should be compatible as far as practical with existing impulse equipment. In addition, the equipment should not be so expensive as to limit its use to those premium installations where the alternative totalizing methods of either paralleled current transformer secondaries or multistator meters cannot be used.

Design Principles

Recent developments, particularly in transistors and logic circuitry, provide the means for the generation and totalization of impulses with static equipment. These devices, along with newly available mechanical components, such as the stepping motor, make possible a system featuring high performance and reliability with low maintenance.

Impulse Generation

Every impulse metering system originates with some type of impulse-sending device driven by the watthour meter or equivalent.

The contact devices heretofore used were difficult to adjust because of the necessity of balancing contact pressure against permissible friction load and because the maintenance needed to provide clean electrical contacts within the meter required continuous attention by experienced personnel.

The concept of a truly modern impulse system has been greatly advanced by the development of the oscillator impulse device shown in Fig. 1. The two principal components of this impulse sender are a pickup head containing a transistorized oscillator appropriately mounted in the meter, and an externally mounted power and relay unit. Fig. 2 shows the schematic circuit diagram of the oscillator and power supply.

The former cams driven by gearing from the meter moving element are replaced by a nonmagnetic segmented disk or vane.

The unit employs a germanium transistor in a special relaxation oscillator circuit. When an open section of the vane moves into position between the coils the oscillator draws a normal current of about 3 milliamperes from the power supply and through the relay coil, closing the normally open contacts. Movement of a solid segment of the vane between the coils reduces the mutual coupling between them so that the oscillator supply current drops to a fraction of a milliampere, thus releasing the relay armature to its normally closed position. The oscillator, which operates at approximately 100 kc is sealed within a molded phenolic head. Single-pole double-throw contacts on the relay provide one complete cycle of operation for each vane point on the disk. This impulse unit is capable of operating at rates up to 300 impulses per minute with negligible friction load on the meter. It may be used with existing impulse equipment of either the 2-wire or 3-

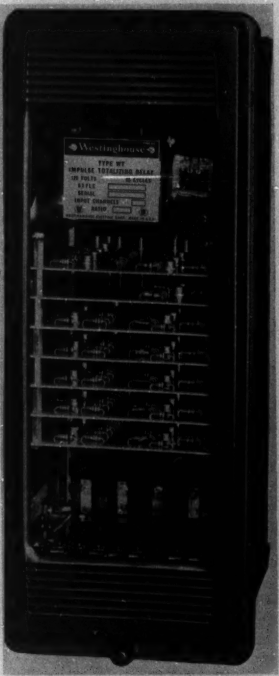


Fig. 3. Impulse totalizing relay

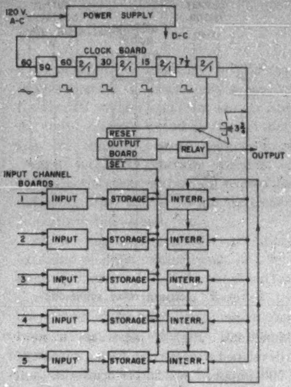


Fig. 4. Block diagram of 5-channel impulse totalizing relay

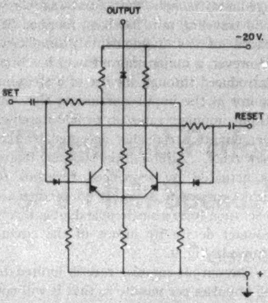


Fig. 5. Typical transistor flip-flop

wire type. Plug-in components facilitate servicing by general field personnel.

Impulse Totalization

The great majority of impulse metering applications involve the totalization of several supply sources or feeders on a single demand receiver. For this prime function it was desirable that the new impulse totalizing relay use static logic or switching components to replace the complex mechanical gadgetry.

Investigation was made into the use of both magnetic core and transistorized noncoincident impulse totalizing systems. The transistorized impulse totalizing relay, shown in Fig. 3, was developed as a result of economic and space considerations.

Units accommodating from 2- to 7-meter input channels are made up by installing a corresponding number of channel boards.

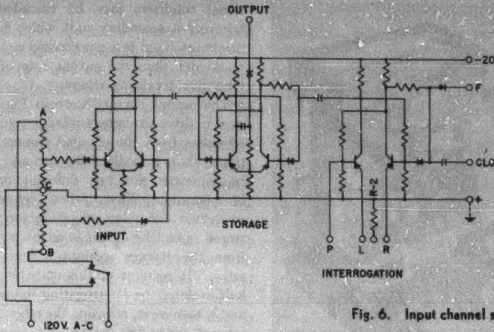


Fig. 6. Input channel schematic

Fig. 4 is a block diagram of a 5-channel impulse totalizing relay. It is seen that there are five channel boards, one for each meter input; an output board and a "clock" board.

The small blocks represent individual function circuits. Nearly all of these circuits consist of two transistors connected as a long-tailed pair with feedback giving a bistable circuit or "flip-flop." A typical unit is shown in Fig. 5 in which capacitive coupling and diode gating have been used for the interconnections between the functional units. This type of circuit may be made to operate with a high degree of reliability in spite of wide variations in component values, supply voltage, and ambient temperature. As long as sufficient drive is supplied to the ON transistor and sufficient bias to the OFF transistor, the circuit will remain in the state in which it has been placed. The diode gating assures reliable switching by steering the pulses to the proper input.

These circuits may be combined in a number of ways to perform many useful functions. For example, the channel board which is shown in Fig. 6 contains three of these to perform the three functions of input, storage, and interrogation. The left-hand pair of transistors comprise the input circuit. The two input terminals are alternately connected to a 120-volt a-c source through a set of contacts in the associated watthour meter.

The input circuit acts as a slave to the watthour meter contacts, reproducing the contact position. With each complete cycle of contact operation the input circuit delivers a "set" pulse to the storage circuit which retains the information until it receives a "reset" pulse from the interrogation circuit. If the storage circuit has been set it will deliver a pulse to the output when it is reset and will then be ready to receive a new pulse. By regularly and sequentially interrogating the storage circuits a string of noncoincident pulses is generated with one output pulse for each input pulse if the input pulse rate does not exceed the interrogation rate.

The interrogation circuits of all channel boards are interconnected to form a ring counter. In a ring counter one of the stages will differ in state from all the others and, by suitably driving the counter, the unique state may be caused to progress from stage to stage. As it progresses, the storage sections are interrogated or "scanned" in sequence. Re-

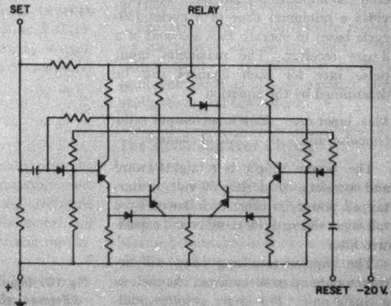


Fig. 7. Output circuit schematic

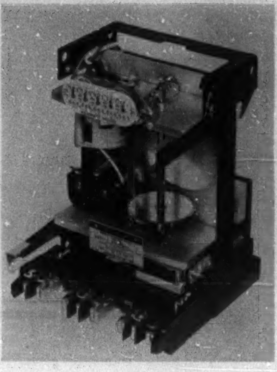


Fig. 8. Impulse storage relay

ferring to Fig. 6 it will be noted that a resistor *R-2* is terminated between the two terminals *L* and *R* of the interrogation circuit. To set up the unique state of one counter stage this resistor is connected to *L* on one board while on all other boards it is connected to *R*. The "ring" is established by connecting from *F* of each circuit to *P* of the following board.

The pulses used to drive the ring counter are obtained from a clock board which contains a pulse squaring stage fed by 60 cycles from the power transformer, and a chain of four binary divider stages. The ultimate output is a square wave with a frequency of 3.75 cycles per second. Higher frequencies are also available from intermediate stages of the divider and are used as will be described later. When the 3.75-cycle frequency is used to drive the ring counter one storage section is interrogated every 16 cycles (60-cycle base).

The response of presently available demand recorders dictated that the maximum impulse output rate should not exceed 225 impulses per minute. This provides a minimum time of 8 cycles (60-cycle base) to operate each solenoid of a 3-wire receiver. The maximum input pulse rate for each channel can be determined by the equation

$$\frac{\text{Max. input rate}}{\text{(Impulse/min.)}} = \frac{225 \times \text{input/output ratio}}{\text{No. of channels}}$$

The power supply is straightforward and consists of a 120/50-volt center-tapped power transformer followed by a full-wave silicon diode rectifier and capacitive filter.

The impulse totalizing relay will accommodate up to seven input channels as a single unit. For larger systems additional totalizers may be cascaded by selecting a secondary unit which has a maximum input rate per channel equal to or greater than the output rate of the preceding totalizing relays.

The output circuit shown in Fig. 7 is used to drive the output relay in response to pulses from the channel storage sections. It is a modified flip-flop and may be connected for two different modes of operation. First, when connected to deliver one relay operation for each stored pulse, the circuit is set by a pulse from the storage sections, closing the relay. It remains in this state for one half cycle of the interrogating frequency and is then reset, releasing the relay, by a squared wave 180 degrees out of phase with the interrogating frequency.

For the second mode of operation the output circuit is set and reset in sequence by pulses from the storage sections. Thus, two input pulses are required to deliver one output impulse, causing the relay to have a 2 to 1 input/output ratio. To maintain an 8-cycle (60-cycle base) relay closure the storage sections are interrogated by the $7^1/_2$-cycle frequency from the clock.

A 4 to 1 input/output ratio is made available by adding an additional binary stage to feed the output circuit. This permits interrogation at 15 cycles for the same chosen output rate.

Impulse Storage

Occasionally need will arise for storing impulses during periods when transmission channels are in use for communication, relaying, or other higher priority

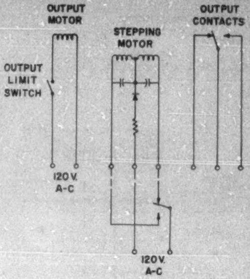


Fig. 9. Storage relay schematic

functions. Fig. 8 illustrates a newly developed storage relay with a capacity of 500 impulses which performs this function and retransmits the stored impulses on demand.

The general principle of previous storage units; namely, that of a storage screw and traveling nut, has been retained because of its fundamental simplicity. However, a major improvement has been introduced through the use of a stepping motor as the impulse actuating element. Input impulses rates up to 500 impulses per minute are readily accepted by this new relay. Although the stepping motor is actuated by sequenced reversals of direct current it is readily adapted to a-c operation from a single-pole double-throw contact device by means of the circuit shown in Fig. 9.

The output impulse rate is limited to 45 impulses per minute so that it will not exceed the maximum continuous input rate of several existing demand receivers.

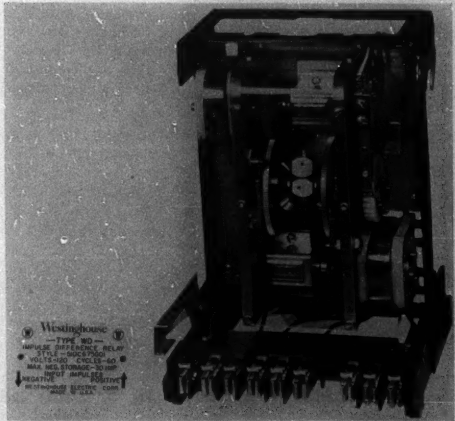


Fig. 10. Impulse difference relay

Net Totalization

In addition to totalizing several impulse sources it is occasionally necessary to subtract a minor load from a larger one. A typical example would be the subtraction of station service power from several totalized generator circuits to obtain net station output. The impulse difference relay in Fig. 10 was developed to perform such a function.

Positive impulses received by the upper left-hand stepping motor result in a forward movement of the differential output shaft. Negative impulses received by the lower right-hand stepping motor cause a backward motion of the output shaft.

The net movement of the differential output shaft is thus proportional to the net difference between the positive and negative impulses.

The output contacts and cam are driven by a pawl responsive to forward movement of the differential but with sufficient freedom of reverse movement to permit the storage of 30 negative impulses.

In this unit also the previously used mechanical differential has been employed because of its basic simplicity. The major improvement in performance comes from the utilization of stepping motors to drive opposing sides of the precision differential. A rate of 500 impulses per minute is acceptable to either the positive or negative inputs.

In applying an impulse difference relay it is imperative that the major, or positive, circuit impulse rate always exceed that of the minor, or negative, circuit. The storage capacity of 30 impulses is provided to allow for brief periods when negative impulses received may temporarily exceed the incoming positive impulses.

Summary

A series of new impulse metering devices is now available which will allow impulse rates to be increased tenfold with a comparable increase in accuracy. The units are adaptable for use with other types of impulse equipment presently in service, or they can be used for completely new high rate installations.

Provisions have been made for even higher impulse rates with minor modifications. For example, the totalizer clock board is equipped with taps for building 1 to 1 or 2 to 1 ratio units with an output rate of 1,800 impulses per minute should such totalizers be desired. Utilization of this potential performance would require demand recorders of much higher speed than any now available. It may be assumed that the output relay of the totalizer would be replaced by a power transistor stage capable of directly driving such recorders. The goal of a completely static totalizing system would then be realized by the utilization of the current change of the oscillator impulse device to drive the input circuit of the totalizing relay.

Design and Performance Characteristics of a High-Speed Wide-Chart Recorder

JOHN C. GARRIGUS
ASSOCIATE MEMBER AIEE

DURING the past decade, the requirements for null balance instruments have greatly changed. Dynamic testing applications have increased the need for accurate, high-speed measuring devices. Specifications such as a span of 50 millivolts and full-scale response time of 5 seconds are supplemented with requirements for a span of 1.0 millivolt and a full-scale response time of less than 0.5 second. Such instruments are important in the fields of rocket or jet engine testing, wind tunnel measurements, automatic machine tool positioning control, and analog-to-digital readout systems using high-speed card punch or print-out operations.

This paper describes the electrical and mechanical considerations required for the design of a high-performance wide-chart recorder. The pertinent characteristics of the input and measuring circuits, amplifier, motor, and feedback network are discussed. The effects of an *R-C* (resistance capacitance) low-pass filter upon the step, ramp, and frequency characteristics are considered.

Principles of Operation

The system consists basically of a potentiometer circuit, d-c to a-c converter, input and isolation transformer, electronic amplifier, servomotor, tachometer and feedback network, and recording mechanism, as shown in Fig. 1.

The potentiometer circuit measures an emf (electromotive force) by comparing it with a reference emf. The "error" or difference signal between the measured and reference emf is converted from d-c to a-c by the "chopper" converter. This signal is then amplified by an input transformer and electronic amplifier to a power level sufficient to drive a servo motor. The motor drives the recording pen mechanism and the contact on the potentiometer slide wire in a direction to decrease the error signal to zero. At balance the position of the contact on the slide wire is a measure of the input emf.

To increase the stability of the system, damping is introduced by means of an a-c induction tachometer coupled to the motor shaft. As the motor drives to balance, the a-c signal from the tachometer is fed to the amplifier through a coupling network. This signal is opposite in phase to the error signal.

The servo motor is geared to the recording mechanism and output shaft. The output shaft is available for auxiliary equipment such as alarms, controls, and analog-to-digital converters.

The Potentiometer Circuit

A typical potentiometer circuit for high-speed recording is shown in Fig. 2(A). This "divider" type of circuit has several advantages over the conventional circuit for low-range and multirange recording. Any thermal voltage generated at the sliding contact is reduced by the factor

---

Paper 60-113, recommended by the AIEE Recording and Controlling Instrumentation Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N.Y., January 31–February 5, 1960. Manuscript submitted November 23, 1959; made available for printing November 27, 1959.

JOHN C. GARRIGUS is with The Bristol Company, Waterbury, Conn.

The author wishes to acknowledge the contributions of D. A. Narducci, Jr., and J. A. Wiencko of the Bristol Company, who have provided editorial and technical assistance, and to H. Brady of Space Technology Laboratories (formerly of The Bristol Company), who was responsible for a portion of the development and design of the instrument.

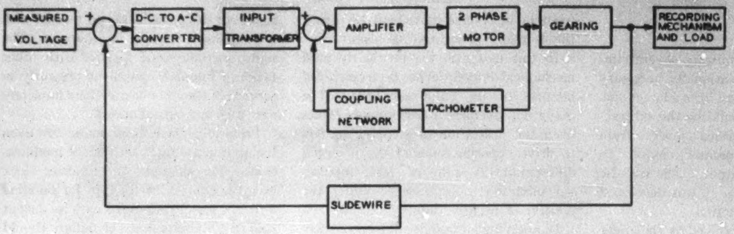


Fig. 1. Basic high-speed recording system

$$\frac{R_{M1}}{R_{M1}+R_{M2}+R_G+R_{SW}+R_D} \qquad (1)$$

on range 1. R_E and $R_F >> R_{SW}$ and R_D. By making this ratio low, contact thermal emf are made negligible.

In a multirange instrument, step or continuous span change is accomplished by changing the amount of the divider resistance included in the input circuit. Fig. 2(A) illustrates the switching arrangement.

In high-speed recording instruments, slide-wire contact bounce seriously affects instrument performance. When the contact leaves the wire, the reference voltage is removed from the balancing circuit. The error voltage then consists of the measured quantity. If the direction of pen travel is toward zero scale position, the response will be overdamped, if toward the maximum scale position, the response will be underdamped. To overcome this action, a capacitor C_0 is added as shown in Fig. 2(B). During normal motion, C_0 maintains the same voltage as E_{SW}. If the contact momentarily opens from points 3 or 4 because of wire irregularity, C_0 maintains voltage E_{SW} until the contact again closes.

The capacitor is small so that normal operation is not affected. However, it is large enough so that the discharge in R_{load}, during a momentary opening of the contact, Fig. 2(C), does not appreciably change the capacitor voltage from the value of E_{SW}. Typical component values are $R_{SW}=150$ ohms, $R_{load}=34{,}000$ ohms, $C_0=1.0$ microfarad.

Detector Circuit

The circuit consists of an electromechanical chopper and an input transformer. The d-c error signal impressed across the chopper input is converted to a square wave a-c signal of power frequency. This signal is then increased by the transformer and fed into the electronic amplifier. The secondary of the transformer is tuned for 60 cps (cycles per second). This provides for minimum phase shift through the input transformer when resistance is inserted in the bridge or input circuit.

Amplifier and Output Circuitry

The amplifier consists basically of four *R-C* coupled voltage-amplifying stages and a power output stage; see Fig. 3. The coupling networks are designed for small phase shift at the carrier frequency and for stability at low frequencies.

The input resistance of the amplifier is 4.7 megohms. The over-all input transformer and amplifier voltage gain is approximately 100×10^6 to detect measured quantities less than 1.0 microvolt. Under maximum stall-torque conditions the output tube delivers 9 watts to the control field of the balance motor.

The control field of the motor is coupled to the amplifier by means of a transformer and series capacitor. The transformer is used to remove the d-c plate current from the control field to minimize motor temperature rise and the capacitor in series tunes the load for maximum power transfer.

For phasing purposes a capacitor is used in series with the reference field of the motor. This capacitor also increases the reference field voltage, thereby increasing the motor speed and torque.

Gain Compensation Network

In multiple-range instruments it is necessary to have constant system gain in order to provide uniform dynamic performance for all ranges. This is usually accomplished by adding resistance in series with the input circuit to maintain the same current sensitivity in the detector circuit. The series resistance increases proportionally with the instrument span. When an *R-C* filter is required in the input circuit, this method cannot be used as the change in resistance affects the filter damping.

Fig. 4 shows a circuit for maintaining constant gain and constant filter resistance in multirange instruments. An auxiliary gain circuit is used to supplement the main gain control, keeping the voltage at the grid of the second-stage constant. This is accomplished by dividing the output voltage of the main control. The instrument is first adjusted for optimum response on the lowest span by R_4. With the span switch indexed to position 2, R_{4A} is adjusted to give the same response. The remaining spans are set similarly. Once all the auxiliary controls are set, the main gain control will affect all ranges similarly.

To keep the filter resistance constant, resistors R_{Y1} and R_{Y2} are added on the lower spans. On each span the resistance must equal

$$\frac{(R_{M1}+R_{M2}+R_{M3})(R_G+R_{SW}+R_D)}{R_{M1}+R_{M2}+R_{M3}+R_G+R_{SW}+R_D} \qquad (2)$$

where R_F and $R_D >> R_{SW}$ and R_D. Then for span 1

$$R_{Y1}=\frac{(R_{M1}+R_{M2}+R_{M3})(R_G+R_{SW}+R_D)}{R_{M1}+R_{M2}+R_{M3}+R_G+R_{SW}+R_D}-R_{M1} \qquad (3)$$

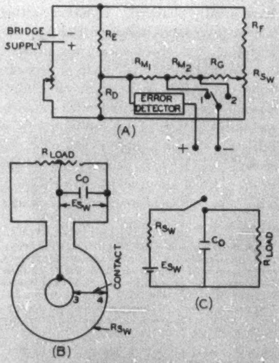


Fig. 2.A—Divider-type bridge circuit. B—Slide-wire portion of bridge circuit diagram showing effect of antibounce capacitor. C—Simplified bridge circuit diagram showing effect of antibounce capacitor

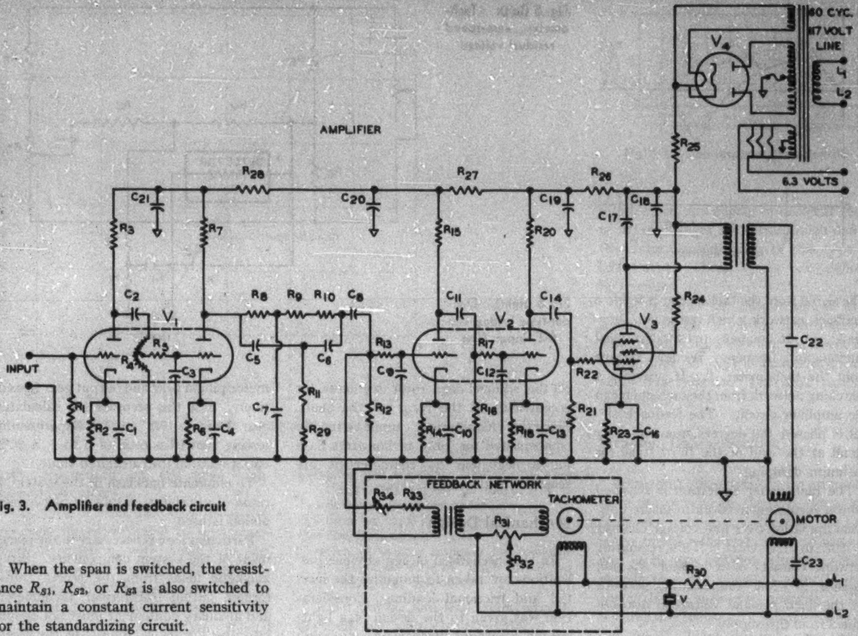

AMPLIFIER
60 CYC.
117 VOLT
LINE
L1
L2
V4
R25
6.3 VOLTS
R28
R27
R26
C21
C20
C19
C18
C17
R3
R7
R15
R20
C2
V1
R8
R9
R10
C8
C11
V2
C14
V3
R24
C22
INPUT
R4
R5
C3
C5
C6
R13
C9
R17
C12
R22
R1
R2
C1
R6
C4
C7
R11
R29
R12
R14
C10
R16
R18
C13
R21
R23
C16
FEEDBACK NETWORK
R34
R33
TACHOMETER
R31
MOTOR
R32
R30
V
C23
L1
L2

Fig. 3. **Amplifier and feedback circuit**

When the span is switched, the resistance R_{S1}, R_{S2}, or R_{S3} is also switched to maintain a constant current sensitivity for the standardizing circuit.

Discrimination Against Line-Frequency Interference

In the basic recorder, line-frequency interference signals that appear as transverse pickup in the input circuitry are converted in the chopper to second and fourth harmonic components. These components are then attenuated in the amplifier by the parallel-T network and *R-C* filter sections at the grids of the second, third, and fourth voltage stages.[1,2] This combination of filters attenuates to the extent that 60-cps spurious signal of 15% of the recorder span does not affect recorder performance.

When the phase of the interference is approximately 90 degrees from line, the major portion of the amplified signal appears in the form

$$\frac{8}{3\pi}\sin 2\omega t+\frac{16}{15\pi}\sin 4\omega t \qquad (4)$$

The second harmonic amplitude is 85% and the fourth harmonic amplitude is 34% of the line-frequency interference.

To reject the second harmonic component, a 120-cps parallel-T filter is designed with a rejection ratio of greater than 500 to 1; the 60 cps is attenuated approximately 5 to 1. This gives a 120-cps to 60-cps rejection ratio of 100 to 1. Tuning is accomplished by adjusting R_9 and R_{29}, Fig. 3, for minimum signal transfer of 120 cps. The fourth harmonic component is attenuated by the *R-C* sections at the grids of the second, third, and fourth stages.

Feedback Circuitry

The damping circuit consists of an a-c induction tachometer and feedback network; see Fig. 3.

The tachometer produces an output voltage proportional to the motor speed.

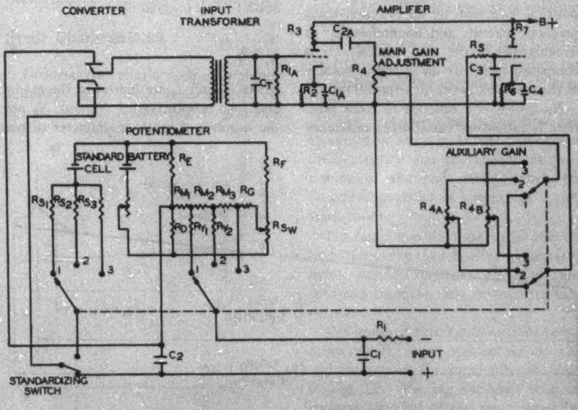

CONVERTER
INPUT
TRANSFORMER
AMPLIFIER
B+
R3
C2A
R7
R5
MAIN GAIN
ADJUSTMENT
C3
CT
R1A
R2
C1A
R4
R6
C4
POTENTIOMETER
STANDARD
CELL
BATTERY
RE
RF
AUXILIARY GAIN
RS1
RS2
RS3
RM1
RM2
RM3
RG
R4A
R4B
RD
RY1
RY2
RSW
1
2
3
R1
C2
C1
−
INPUT
+
STANDARDIZING
SWITCH

Fig. 4. **Gain compensation for multiple-range instruments**

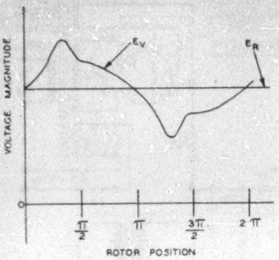


Fig. 5 (left). Tachometer zero-speed residual voltage

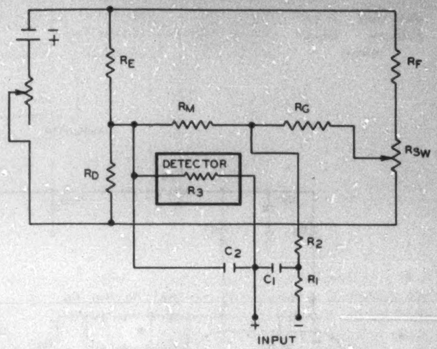


Fig. 6 (right). Divider-type bridge with R-C input filter

The signal from the tachometer is fed to a feedback network which serves two functions: 1. It compensates for the fixed fundamental frequency residual voltage from the tachometer. 2. It provides a matching network from the tachometer to the amplifier circuit. The feedback signal is phased 180 degrees from the error signal at the grid of the third stage for optimum damping.

The tachometer excitation is supplied from a regulating network to insure optimum operation over line voltage changes of 107 to 127 volts.[1] This regulating network consists of a resistor R_{30} and varistor V. The compensating network for the fixed residual voltage is also obtained from this supply.

A damping control R_{34} is provided for adjusting the dynamic performance of the system as desired.

In addition to the tachometer output voltage, which is proportional to the shaft speed, two residual voltages of fundamental frequency exist at zero speed. These voltages E_R and E_V, shown in Fig. 5, affect the damping characteristics.

E_R, which is independent of rotor position, is caused by an unsymmetrical magnetic circuit, and nonuniformity or irregularities of the pole surfaces. It is compensated for by an opposing voltage in the feedback network, adjusted by R_{31}.

E_V, which is a function of rotor position, is caused by variation in resistance of the squirrel-cage rotor segments, or eccentricity of the rotor on the shaft. Since this characteristic cannot be readily compensated for, only tachometers having a negligible E_V characteristic are selected for this use.

Mechanical Design

In the mechanical design obvious precautions are taken to minimize the inertial and frictional loading. Consideration was given to the design of a lightweight recording mechanism, minimum inertia of the output gear, and use of ball-bearing pulleys for the recording mechanism.

To insure adequate speed and acceleration without sacrificing output torque, optimum gearing between motor pinion and output gear is necessary.[3] Assuming negligible motor and load friction, the optimum gear ratio for maximum acceleration to a step signal is

$$\frac{N_2}{N_1}=\sqrt{\frac{J_2}{J_1}} \tag{5}$$

where J_1 and J_2 are inertia of the motor and load respectively; N_1 and N_2 are the number of teeth or diameter of the motor pinion gear and output gear respectively. For this recorder the calculated gear ratio is 7.2 to 1. Experimental testing showed a ratio of 6 to 1, a 20% variation from the calculated value.

To eliminate backlash in the system, a cause of instability, a split, spring-loaded pinion is used.

Particular care is necessary in the placement of the system components. Stray magnetic fields from the power transformer, output transformer, tachometer, and auxiliary equipment must not couple into the low-level circuits, or the result will be offset, dead band, and instability.

Input Circuit Filtering

In some applications it is necessary to suppress large interference voltages that are superposed on the measured quantity. A number of filter configurations have been designed consisting of resistive, capacitive, and inductive components to form low-pass or band-rejection networks. Although attenuating the unwanted frequencies, a filter will affect the instrument performance in one or more of the following ways: 1. an increase of system time lag; 2. a decrease of dynamic performance; 3. an effective lowering span of the recorder by attentuating the d-c error signal by the ratio of

$$\frac{R_3}{R_S+R_F+R_M+R_3} \tag{6}$$

where R_3 is the detector resistance, R_S is the external circuit resistance, R_F is the sum of the filter resistances, and R_M is the bridge resistance.

Resistive and capacitive components are preferred to inductive components. Although the inductive elements attenuate low frequencies with less time lag, they require tight tolerances and considerable shielding for optimum perform-

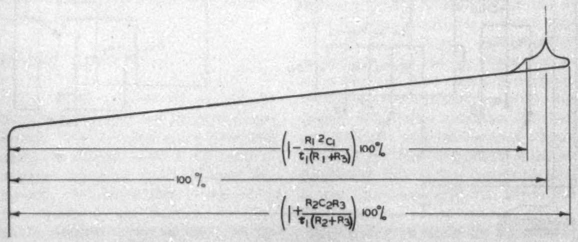

$\left(1-\frac{R_1{}^2C_1}{t_1(R_1+R_3)}\right)100\%$
100 %
$\left(1+\frac{R_2C_2R_3}{t_1(R_2+R_3)}\right)100\%$

Fig. 7. Response for a 100% step input signal illustrating lag and lead characteristics

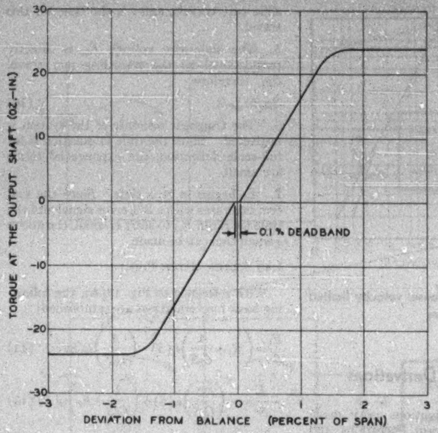


Fig. 8. Static torque characteristic

ance, particularly in low-range high-speed instruments.

One type of filter used consists of *R-C* low-pass stages arranged to give maximum attenuation, with small time lag, and does not affect the recorder performance. Such a filter is shown in Fig. 6. It consists of lag and lead sections. R_1 and C_1 comprise the lag section; R_2 and C_2 comprise the lead section.

The lag section makes the recorder damp or "hitch" at approximately

$$\left(1-\frac{R_1^2C_1}{t_1(R_1+R_3)}\right)100\% \tag{7}$$

of a full-scale step response signal, as shown in the Appendix and illustrated in Fig. 7.

The lead section makes the recorder overshoot by approximately

$$\left(1+\frac{R_2R_3C_2}{t_1(R_2+R_3)}\right)100\% \tag{8}$$

of a full-scale step response signal, where

R_1 = filter lag resistance
R_2 = filter lead resistance (bridge resistance $R_M << R_2$ and can be neglected)
R_3 = detector circuit resistance
C_1 = filter lag capacitance
C_2 = filter lead capacitance
t_1 = slope of system response for full-scale response without filter (time for 100% step response)

When the lag and lead characteristics are combined, a filter is provided that will not affect the recorder damping characteristics. In the Appendix it is shown that such a filter must satisfy the equation

$$(R_1+R_2)R_3\left(\frac{C_2}{C_1}\right)=R_1^2 \tag{9}$$

From equation 9 it is noted only the ratio of C_2/C_1 is important for optimum dynamic response, and not the magnitudes.

Static Characteristics

Performance criteria such as accuracy linearity, dead band, repeatability, rejection of transverse and longitudinal interference, and effects of ambient temperature changes are terms which define the static characteristics of this type of instrument. These terms are defined in reference 1.

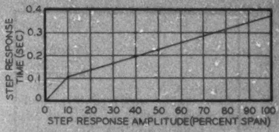


Fig. 10. Step response characteristic

Fig. 8 shows the output torque as a function of the deviation from balance. The system saturates when the deviation is 2%, and provides a maximum torque of 24 ounce-inches. The dead band is less than 0.1%.

Dynamic Characteristics

Important characteristics for an instrument of this type are the step, ramp, and frequency responses. Step response time is the time for the end device to come to rest in its new position after an abrupt change to a new constant value has occurred in the measured quantity.[1] Fig. 9 shows a response to a full-scale step input signal. Total time to balance (where balance is defined as coming to rest within the accuracy rating of the instrument) is approximately 0.37 second, of which 0.06 second is required for acceleration and 0.06 second for deceleration. The response time for step-input signals from zero to 100% is shown in Fig. 10. The slope from zero to 10% indicates the region in which the system is accelerating. Above 10% the system is velocity-limited.

By proper use of the damping control, inertial loads up to 25 ounce-inches can be applied[2] to the output shaft without affecting the step response time.

Ramp response time is the time lag, expressed in seconds, between the measured quantity and the equivalent positioning of the end device, when the measured quantity is varying at a constant rate.[1]

The ramp response time can be affected by variations in the following parameters: motor voltage constant, amplifier gain, damping constant, and input filter time constant.

The graph of time lag in seconds versus ramp velocity in per cent per second with and without an *R-C* input filter is shown in Fig. 11. The lag decreases with increasing velocity of the ramp input because of motor voltage constant, ampli-

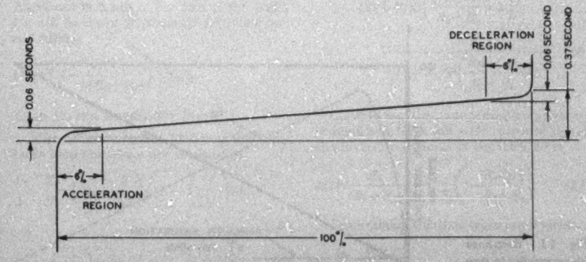


Fig. 9. Response to a 100% signal illustrating acceleration and deceleration region

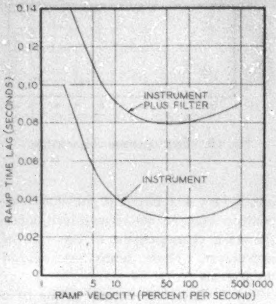


Fig. 11. Ramp response time

fier gain, and damping constant, until the saturation velocity is reached. Above the saturation velocity the system cannot follow as the amplifier and motor saturate, and the time lag increases.[5]

The introduction of an input filter increases the response time by the lag section filter time constant R_1C_1. From Fig. 11, the experimental results indicate an increase of about 0.05 second.

Frequency response is the dynamic response of an instrument when the measured quantity is varied sinusoidally over a band of frequencies.[1] Frequency response curves for a 100% peak-to-peak amplitude are shown in Fig. 12. These are curves of the system frequency response characteristics with the *R-C* filter A and without the *R-C* filter B. For any other input greater than 10% peak to peak, the response can be approximated by drawing a horizontal line corresponding to the input magnitude and extending to the minus 2 slope line. With input signals less than 10%, the system is not velocity-limited and the curve is not valid.[5] (An input of 25% peak to peak is illustrated by the dashed line.) At approximately 20 cps the attenuation curve changes from a minus 2 slope to a minus 3 slope due to the attenuation characteristics of the filter.

Conclusions

The principles of operation of a high-speed industrial recorder have been discussed. By considering the characteristics of the *R-C* input filter and potentiometer circuit, amplifier, motor, and feedback networks a high-performance instrument was developed. Instruments of this type are being applied to many industrial applications where high torque, sensitivity, and speed are required.

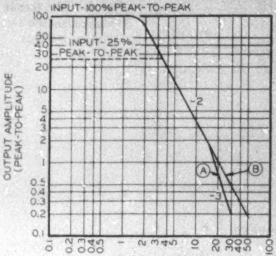


Fig. 12. Frequency response: velocity limited system

Appendix. Derivations

In the following derivations use is made of the Laplace transform operations and pairs for solving the equations.[4]

Lag and Lead Filter Sections

R_1 and C_1 have been described as constituting the lag section of the filter. Lag in this sense means a damping, or "hitch" in the response characteristic. This type of response occurs when the filter causes the error voltage to become zero before the true balance point is reached.

The lead section, consisting of R_2 and C_2, causes the opposite effect, making the instrument overshoot. This response occurs when the error voltage becomes zero after the true balance point is reached.

In the derivation of the equations in this Appendix, the following assumptions are made; see Fig. 13:

1. The measured quantity E_1 is a step with magnitude equal to the span of the instrument.

2. The system is critically damped without the input filter, i.e., the pen moves at a constant rate across the scale and stops abruptly at balance.

3. The amplifier saturates with a 2% static error signal.

4. The system is velocity limited, causing the slide-wire voltage E_2 to equal Kt where $K\alpha S_W/t$. S_W is the distance which the slide wire travels, and t is the time for this travel.

5. The slide-wire voltage E_2, is directly proportional to the recording pen travel S_P; therefore

$$S_W/t\alpha S_P/t \tag{10}$$

6. The transient response of the system is neglected. Since the time in question is for full-scale deflection, the exponential terms are small.

7. E_2 begins in zero time. Since the system saturates with a 2% error signal and the time for ramp E_2, to start is small, the above assumption can be made.

LAG FILTER HITCH POINT

With reference to Fig. 14(A), the following basic loop equations are established:

$$\frac{E_1}{S}=\left(R_1+\frac{1}{C_1S}\right)i_1(S)-\left(\frac{1}{C_1S}\right)i_2(S) \tag{11}$$

$$-\frac{K}{S^2}=-\left(\frac{1}{C_1S}\right)i_1(S)+\left(\frac{1}{C_1S}+R_2\right)i_2(S) \tag{12}$$

where E_1/S is the transform for a step signal of magnitude E_1, and K/S^2 is the transform for a ramp signal. Kt is generated by the repositioning of the slide wire. The sign is negative because Kt is opposing E_1.

Solving for $i_2(S)$ yields

$$i_2(S)=\frac{E_1}{(R_1+R_2)\left(\frac{R_1R_2C_1S}{R_1+R_2}+1\right)S}-\frac{K(R_1C_1S+1)}{(R_1+R_2)\left(\frac{R_1R_2C_1S}{R_1+R_2}+1\right)S^2} \tag{13}$$

Considering the exponential components insignificant for a full scale response signal, the inverse transform of equation 13 is

$$i_2(t)=\frac{E_1}{R_1+R_2}-\frac{K}{R_1+R_2}\left(R_1C_1-\frac{R_1R_2C_1}{R_1+R_2}+t\right) \tag{14}$$

Determining the time when $i_2=0$ gives

$$t=\frac{E_1}{K}+\frac{R_1R_2C_1}{R_1+R_2}-R_1C_1 \tag{15}$$

Since E_1 = a step signal equal to E_0 (full span voltage) and K = the slope of the recording pen or slide-wire voltage during maximum speed across the chart equal to

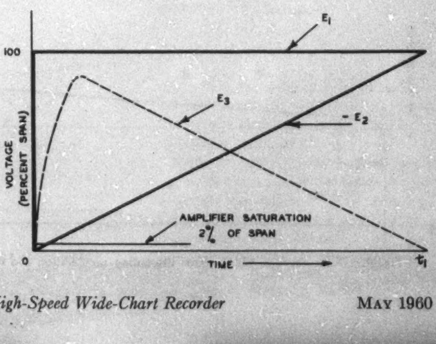


Fig. 13. Response characteristics to step input signal

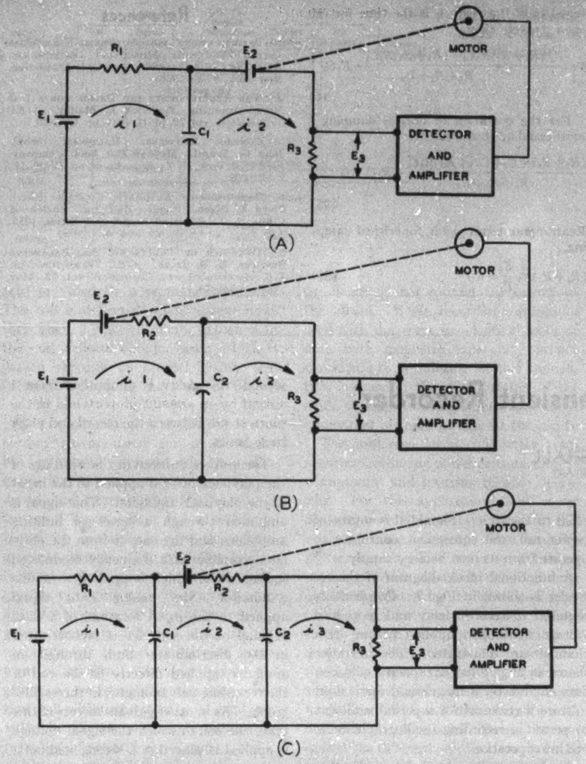


Fig. 14. A—Simplified schematic of R-C lag filter and bridge network. B—Simplified schematic of R-C lead filter and bridge network. C—Simplified schematic of combined R-C lag and lead filter and bridge network

E_0/t_1, where t_1 is the time for full scale travel. Then

$$t=t_1-\left(\frac{R_1^2C_1}{R_1+R_3}\right) \tag{16}$$

Expressed in terms of per cent of full scale, t/t_1 will be the point where the recording pen will hitch:

$$t/t_1=\left(1-\frac{R_1^2C_1}{t_1(R_1+R_3)}\right)100\% \tag{17}$$

LEAD FILTER OVERSHOOT POINT

With reference to Fig. 14(B) the following basis loop equations are established:

$$\frac{E_1}{S}-\frac{K}{S^2}=\left(R_2+\frac{1}{C_2S}\right)i_1(S)-\left(\frac{1}{C_2S}\right)i_2(S) \tag{18}$$

$$0=-\left(\frac{1}{C_2S}\right)i_1(S)+\left(R_3+\frac{1}{C_2S}\right)i_2(S) \tag{19}$$

where E_1 and K are the same as stated for the lag filter hitch point. Solving for $i_2(S)$ yields

$$i_2(S)=\frac{E_1}{(R_2+R_3)S\left(\frac{R_2R_3C_2S}{R_2+R_3}+1\right)}-\frac{K}{(R_2+R_3)S^2\left(\frac{R_2R_3C_2S}{R_2+R_3}+1\right)} \tag{20}$$

Considering the exponential components insignificant for a full scale response signal, the inverse transform of equation 20 is

$$i_2(t)=\frac{E_1}{R_2+R_3}-\frac{K}{R_2+R_3}\left(t-\frac{R_2R_3C_2}{R_2+R_3}\right) \tag{21}$$

Determining the time when $i_2=0$ yields

$$t=\frac{E_1}{K}+\frac{R_2R_3C_2}{R_2+R_3} \tag{22}$$

With the same conditions as stated in equation 15

$$t=t_1+\frac{R_2R_3C_2}{R_2+R_3} \tag{23}$$

Expressed in terms of per cent of full scale, t/t_1 will be the maximum point the recording pen will overshoot:

$$t/t_1=\left(1+\frac{R_2R_3C_2}{t_1(R_2+R_3)}\right)100\% \tag{24}$$

LAG AND LEAD CHARACTERISTICS COMBINED

By combining the lag and lead characteristics, a filter can be designed that will not affect the damping characteristics. With reference to Fig. 14(C) the following loop equations are established:

$$\mathcal{L}[E_1(t)]=\left(R_1+\frac{1}{C_1S}\right)i_1(S)-\left(\frac{1}{C_1S}\right)i_2(S) \tag{25}$$

$$-\mathcal{L}[E_2(t)]=-\left(\frac{1}{C_1S}\right)i_1(S)+\left(R_2+\frac{1}{C_1S}+\frac{1}{C_2S}\right)i_2(S)-\left(\frac{1}{C_2S}\right)i_3(S) \tag{26}$$

$$0=-\left(\frac{1}{C_2S}\right)i_2(S)+\left(R_3+\frac{1}{C_2S}\right)i_3(S) \tag{27}$$

where $E_1(t)$ and $E_2(t)$ are the exciting voltages.

Solving for i_3, clearing factors and combining terms, yields the general filter equation

$$i_3(S)=\frac{\mathcal{L}[E_1(t)]-\mathcal{L}[E_2(t)](R_1C_1S+1)}{(R_1R_2R_3C_1C_2)S^2+[(R_1+R_2)R_3C_2+(R_2+R_3)R_1C_1]S+R_1+R_2+R_3} \tag{28}$$

Let

$X=R_1R_2R_3C_1C_2$
$Y=(R_1+R_2)\ R_3C_2+(R_2+R_3)\ R_1C_1$
$Z=R_1+R_2+R_3$

$$i_3(S)=\frac{\mathcal{L}[E_1(t)]-\mathcal{L}[E_2(t)](R_1C_1S+1)}{XS^2+YS+Z} \tag{29}$$

In time-constant form

$$i_3(S)=\frac{\mathcal{L}[E_1(t)]-\mathcal{L}[E_2(t)](R_1C_1S+1)}{(T_1S+1)(T_2S+1)(R_1+R_2+R_3)} \tag{30}$$

where

$$T_1T_2=\frac{X}{Z}$$

$$T_1+T_2=\frac{Y}{Z}$$

For a step input $\mathcal{L}[E_1(t)]=E_1/S$ and $\mathcal{L}[E_2(t)]=K/S^2$

$$i_3(S)=\frac{E_1}{S(T_1S+1)(T_2S+1)(R_1+R_2+R_3)}-\frac{K(R_1C_1S+1)}{S^2(T_1S+1)(T_2S+1)(R_1+R_2+R_3)} \tag{31}$$

Again considering the exponential components insignificant for a full scale response signal, the inverse transform of equation 31 is

$$i_3(t)=\frac{E_1}{R_1+R_2+R_3}-\frac{K}{R_1+R_2+R_3}\times(t-(T_1+T_2)+R_1C_1) \qquad (32)$$

Determining the time when $i_3=0$ and solving for t yields

$$t=\frac{E_1}{K}+\frac{(R_1+R_2)R_3C_2+(R_2+R_3)R_1C_1}{R_1+R_2+R_3}-R_1C_1 \qquad (33)$$

Since E_1=a step signal equal to E_0(full span voltage), and K=the slope of recording pen travel or slide-wire voltage during maximum speed across the chart, and is equal to E_0/t_1, where t_1 is the time for full span travel, then

$$t=t_1+\frac{(R_1+R_2)R_3C_2+(R_2+R_3)R_1C_1}{R_1+R_2+R_3}-R_1C_1 \qquad (34)$$

For the condition of critical damping t must equal t_1; therefore

$$\frac{(R_1+R_2)R_3C_2+(R_2+R_3)R_1C_1}{R_1+R_2+R_3}-R_1C_1=0 \qquad (35)$$

Rearranging terms yields, for critical damping,

$$(R_1+R_2)R_3\frac{C_2}{C_1}=R_1^2 \qquad (36)$$

References

1. SPECIFICATIONS FOR AUTOMATIC-NULL BALANCING ELECTRICAL MEASURING INSTRUMENTS. *ASA C39.4-1956*, American Standards Association, New York, N. Y., 1956.

2. THE REQUIREMENTS AND DESIGN FOR A D-C NULL DETECTOR, **Frederick L. Maltby.** *AIEE Transactions*. vol. 70, pt. II, 1951, pp. 1876-83.

3. CONTROL ENGINEERS' HANDBOOK (book), **John G. Truxal.** McGraw-Hill Book Company, Inc., New York, N. Y., first edition, pp. 12-56, 57, 3-31, 13-8.

4. PRINCIPLES OF AUTOMATIC CONTROL (book), **Floyd E. Nixon.** Prentice-Hall, Inc., Englewood Cliffs, N. J., Electrical Engineering Series, 1953, App. 6.

5. DYNAMICS OF ELECTRONIC SELF-BALANCING SYSTEMS, **G. R. Jacob.** *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 76, May 1957, pp. 166-74.

A Power-Line Transient Recorder

C. H. HOSHALL
NONMEMBER AIEE

THOSE WHO have been associated with the industry during recent years have witnessed a most progressive expansion and improvement of electric power generation and distribution facilities. This growth has been accompanied and perhaps preceded in many instances, by improvements in instrumentation techniques and equipment. The recorder described here has capabilities which exceed those of similar recorders in general use. It provides a means of recording waveforms which exist in a power system immediately prior to, during, and after the occurrence of a fault or other disturbance without continuous consumption of photographic recording medium. This instrument will record current and voltage waveforms, or other waveforms if desired, as long as required after the disturbance occurs.

General Description

The field-test model of this equipment, with front doors removed, is shown in Fig. 1. The unit has been made entirely self-contained so that it can be moved about and readily adapted to field-test conditions. Under normal operating conditions, 117-volt 60-cycle power is supplied to the unit. Should this source of power fail, the equipment continues to operate from its own battery supply.

A functional block diagram of the recorder is shown in Fig. 2. Components required to record, delay and reproduce voltage waveforms applied to one data channel are illustrated. The recorder shown in Fig. 1 has a capacity of seven data channels. A magnetic drum is used to store waveforms for a period sufficient to permit a recording oscillograph to be put into operation.

As shown in Fig. 2, the input signal is applied to a voltage-controlled oscillator. The output of this oscillator is frequency modulated by the input signal, and is applied to an amplifier which supplies sufficient power to drive a recording head. Magnetic particles which have been applied to the storage surface of the drum are influenced by the magnetic field produced by current through the winding of the head; and data is recorded on the drum in a manner analogous to that in which saturation recording is accomplished in conventional frequency modulation magnetic tape instrumentation systems.

Slightly less than 1 second after any point on the periphery of the drum passes under the record head, it passes under the gap of the playback head. The recorded pattern is then "erased" as it passes under a permanent-magnet erase head which is located between the playback and record heads. Thus, a delay of slightly less than 1 second is provided. Since the drum makes one revolution each second, this delay is diminished from 1 second by an amount proportional to the shortest arc between the record and playback heads.

The voltage induced in the windings of the playback head is applied to the input of the playback amplifier. This signal is amplified through a series of limiting amplifiers, and the output from the playback amplifier is a frequency modulated square wave, which is applied to a discriminator. The analog data signal applied to the input terminal of a data channel of the recorder is reconstituted in the discriminator and, through an amplifier, applied directly to the coil of the recording galvanometer in the oscillograph. As in other galvanometers of this type, the coil to which the signal voltage is applied is placed in a strong magnetic

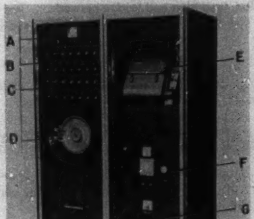


Fig. 1. Fault recorder, field-test model, doors removed

A—Voltage regulator, setup instrumentation
B—Record modules
C—Playback modules
D—Magnetic drum
E—Recording oscillograph
F—Battery charger
G—Battery supply

Paper 60-112, recommended by the AIEE Recording and Controlling Instrumentation Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., January 31-February 5, 1960. Manuscript submitted November 23, 1959; made available for printing January 18, 1960.

C. H. HOSHALL is with the Minneapolis-Honeywell Regulator Company, Beltsville, Md.

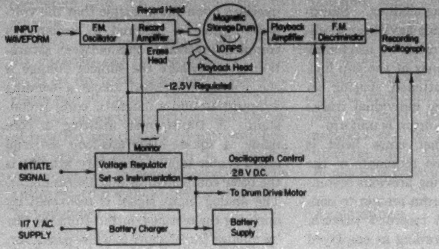


Fig. 2. Functional block diagram of fault recorder

field provided by a permanent magnet. The coil is deflected when current flows through it. A small mirror attached to the coil deflects a light beam which is passed through an optical system and brought to focus in the plane of the recording film.

The signal voltage is applied to the recording galvanometer coil at all times, but no record is made on the oscillograph film because lamp brilliance is reduced and because the film is not in motion. When a fault occurs and is detected by initiating relays, the oscillograph is started. The recording lamp is brought to full brilliance, and the photographic recording paper is drawn past an aperture in which the light beam from the galvanometer is deflected along an axis perpendicular to the direction of movement of the film. The recording oscillograph can be brought to operating condition well within the storage interval provided by the drum, thereby permitting recording of prefault waveforms.

Also shown in Fig. 2 is the instrumentation unit. This unit contains a voltage regulator which provides regulated -12.5 volts from which all the electronic circuits in the recorder operate. Also contained in the instrumentation unit are a reference oscillator, reference channel recording and playback module, and circuits with which operation of the recorder can be monitored once initial calibrations have been made.

A 120-cps (cycles per second) sinusoidal reference signal is recorded on the drum, played back, and recorded on the oscillograph in exactly the same manner as the data signals. This reference signal aids in the analysis of the oscillograph records.

Design Considerations

The very important requirement that the unit function reliably for extended periods of time without adjustment or service has been met by using a magnetic drum as the storage medium. There is no wear of the recording medium because the heads do not contact the surface of the drum. While recording techniques used with the drum are identical to those used with magnetic tape, tape is not considered to be suitable for this application because of the deterioration of the oxide coating resulting from wear caused by contact of the tape against the heads.

The fault recorder is completely transistorized, reducing power consumption to a minimum and insuring reliable operation. For this application, low power consumption is important because this equipment must operate from standby power sources which may be of limited capacity. Also, transistorized circuits are ideally suited for modular packaging utilizing printed circuit cards. These cards provide for maximum consistency from unit to unit. The modular packaging used in this equipment permits simple and rapid exchange of plug-in modules in the event that circuits with different characteristics should be required for different applications, or in the event of circuit failure.

Since the field-test equipment described here has a capacity of seven data channels and one reference channel, each head used has a total of eight tracks. All tracks of the record and playback heads are assembled and cast as integral units. This simplifies adjustment and alignment of the heads.

Wide-band FM (frequency modulation) recording is used because it is the only recording technique readily adaptable for this application that can provide the required amplitude linearity. This is necessary so that the amplitude of the recorded waveforms can be accurately scaled, and so that the recorder introduces only negligible distortion in waveforms which it preserves. Another important advantage of FM recording is that irregularities in the level of the signals played back from the drum have less influence on an FM recording system than on other types of systems. Irregularities in playback signal level may be caused by slight imperfections in the drum surface or by nonhomogeneity of the magnetic dispersion applied to the drum.

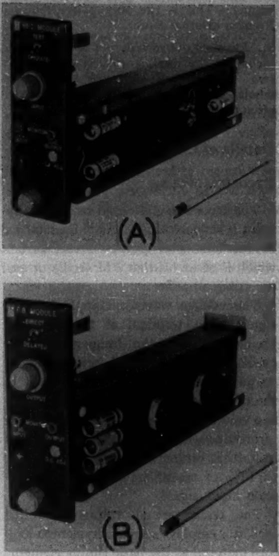


Fig. 3. Fault recorder electronic units

A—Record module
B—Playback module

General Specifications

Performance specifications of the field-test equipment, in addition to those which have already been mentioned, are as follows:

1. Data frequency response: The response of the entire unit from input terminal to oscillograph record is flat to within ±10% from 45 cps to 500 cps.
2. Amplitude accuracy (linearity) of data channels at all frequencies in the data pass band: The amplitude response from input terminal to oscillograph record is linear to within ±3% of full-scale deflection.
3. Signal input voltage: 2.0-volt rms causes full-scale deviation of the recording oscillator. An adjustable input attenuator is provided.
4. Oscillator input impedance: approximately 1,000 ohms.
5. Maximum oscillograph record amplitude: 4 inches, peak-to-peak.
6. Timing marker frequency accuracy: ±2%.
7. Power requirements: The unit operates from a 24-volt storage-battery supply. The battery charger requires 117-volt a-c

supply. This is the only power required from an external source.

8. Physical dimensions: The entire unit, including the oscillograph, is mounted in two standard 19-inch-wide Radio-Electronics-Television Manufacturers Association relay racks, 5 feet high.

Details of Circuit Operation

Record Module

The record module used in this equipment is shown in Fig. 3(A). Each record module houses one printed circuit card on which is assembled an FM oscillator and recording amplifier.

The voltage waveform applied to the signal input terminal of the recording oscillator causes the frequency of the oscillator to be deviated from its center frequency by an amount proportional to the deviation of the signal from its average value. The center frequency of the recording oscillator is 3 kc, and deviation from this center frequency is ±1.2 kc (±40% of center frequency) with full-scale input signal.

The transistorized voltage-controlled oscillator employs a square-hysteresis-loop core material as the basic timing element. The oscillator is well suited for this application because it has a very linear frequency-versus-voltage characteristic, is simple, and is stable despite temperature variations.

Front panel controls, shown in Fig. 3(A), include an input selector switch (for selection of either test or data signal), an input attenuator control, and a control for adjusting the center frequency of the oscillator. Monitor jacks facilitate monitoring of waveforms at selected points in the circuit.

Magnetic Storage Drum

The magnetic storage drum used in this recorder is 10 inches in diameter and $1^3/_4$ inches wide. It is an accurately machined metallic drum made of a nonmagnetic material. A dispersion coating, which has magnetic properties similar to the oxide used on magnetic tape, is applied to the prepared surface of the drum. Since it is difficult to apply a sufficiently smooth coating of this dispersion by direct application, an excess of oxide is applied and the surface is machined to remove irregularities. The thickness of the finished coat is approximately 0.001 inch.

The drum is driven by a governor-controlled, d-c motor which is operated from the fault recorder storage-battery supply. Belts and pulleys are used to provide a drum speed of 1 revolution per second.

Record and playback head gaps are spaced approximately 0.001 inch from the surface of the recording medium. Standard magnetic tape heads have been modified for use with the drum by machining the heads along the apex at the gap line. With the heads so flattened, spacing between the gaps of the individual tracks and the surface of the drum is uniform.

A permanent-magnet erase head is used to insure that no remnants of the patterns recorded during previous revolutions of the storage drum remain to contaminate the desired recorded pattern. In FM saturation recording as employed in magnetic-tape systems, the recording head is in direct contact with the recording medium. A separate erase head is not usually required because the magnetic pattern is completely re-established each time any portion of the tape passes over the record head gap. Because it is considerably more difficult to completely saturate the magnetic medium when it does not contact the record head, a separate erase head has been supplied.

Playback Module

The playback module used in this equipment is shown in Fig. 3(B). Two printed circuit cards are mounted in each playback module housing. One card is the playback amplifier; the other is a discriminator and amplifier.

The playback head output is approximately 3 millivolts rms with the head unloaded and with an unmodulated 3-kc carrier signal recorded on the drum. The playback amplifier output, when unloaded, is a square wave in which the transitions occur at the instant the input voltage passes through zero. In addition to amplifying and limiting the signal from the drum, the playback amplifier supplies sufficient power to operate the discriminator, which is connected to its output.

The discriminator circuit used in this instrument employs a magnetic core as the frequency-sensitive element, and operates upon the principle that the volt-second integral required to accomplish complete reversal of flux between saturation levels (where the subject voltage is the voltage impressed across a winding wrapped around a magnetic core) is constant. A transformer utilizing a core made of square-hysteresis-loop material and a center-tapped secondary winding is the basic component of this discriminator. The analog input signal is recovered by applying the rectified output from the discriminator transformer to a low-pass filter. The cutoff frequency of this filter is approximately 550 cps. Carrier and higher frequencies are attenuated, and data frequency signals are passed with a minimum of attenuation. The analog signal output from the filter is applied to a feedback-stabilized transistor amplifier which supplies sufficient power to operate the oscillograph galvanometers. The output signal can be attenuated, if desired, by use of an L-pad output control mounted on the front panel of the playback module.

Front panel controls are provided for circuit adjustment and monitoring. A small portion of the recording module output can be applied to the input of the playback amplifier by means of a toggle switch on the playback module. In this manner, the magnetic storage drum is bypassed and adjustments or tests can be made when the drum is not in operation.

Instrumentation Unit

Other circuits essential to operation of the system are contained in the instrumentation unit. The electronic units of the recorder must be supplied from a low-impedance power source, and the supply voltage must be constant despite changes in battery supply voltage caused by variations in battery condition and load.

The instrumentation unit contains all circuits required to set up the recording

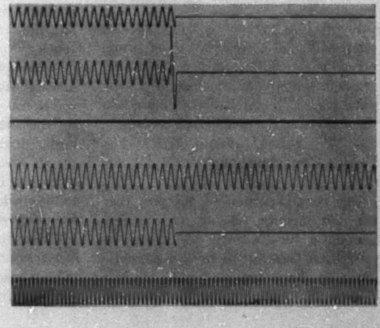


Fig. 4. Laboratory record

oscillators and discriminators, and to monitor battery and regulated supply voltages. Selected points in all recording channels can be conveniently monitored on the front panel meter, and at a single pair of monitor jacks on the front panel of the instrumentation unit by use of function and channel selector switches. Sufficient instrumentation is provided within the unit to permit operation of the entire recorder without the use of external test equipment.

The reference signal recording and playback module located in the instrumentation unit consists of three printed circuit cards which are, except for small differences in external wiring, identical to the FM data-channel cards.

OSCILLOGRAPH

The oscillograph is a standard unit that uses fluid-damped galvanometers flat to within ±5% from 0–600 cps. To insure long lamp life, the recording lamp is operated continuously at reduced brilliance, and is brought up to full brilliance upon the advent of a fault. The oscillograph has a capacity of 200 feet of photographic paper which can be run at any rate between 144 and 0.031 inches per second. The amount of film run out upon each initiation can be preset with a front panel control. Record numbers can be photographed directly upon the film. The quality of the record produced is illustrated in Fig. 4.

Applications

The field-test equipment discussed here is packaged in two racks. However, the delay unit is a distinct and separate unit and, as such, can be operated entirely independently of the equipment mounted in the companion rack. The storage drum and associated equipment can be used for applications other than for recording power-line transients and for fault recorder applications where operation of the equipment from existing station battery supplies is desirable. Any oscillograph can be used which uses galvanometers with sensitivities that will provide the deflection desired, so long as it can be brought to operating condition within 1 second so that prefault information can be recorded.

This recorder can be of considerable value in experimental and developmental work as well as in operational installations. When incorporated into operating systems, it can be used to record voltage or current waveforms at the most advantageous points. When used with suitable shunts and transformers, phase currents and voltages, and residual currents can be recorded. Under-voltage, over-voltage, or over-current relays can be placed in the system to initiate action of the recording oscillograph.

Through analysis of permanent records made with this recorder, location of faults is made easier and the sequence of operation of protective relays can be determined. This recorder can also be used during staged tests to determine if protective devices are properly adjusted and are in good operating condition.

Dial-Selected Industrial Television for Ticket Reservation Facilities

J. W. ALINSKY
NONMEMBER AIEE

ABOUT 6 years ago Pennsylvania Railroad found that an improvement in the method of handling ticket sales and reservations was necessary in order to decrease the time and effort required to process ticket sales. In the following years, they designed and built new ticket facilities at Philadelphia and Pittsburgh, Pa., experimenting with new ideas and new equipment until they had a system operating which would meet the transportation demands of these cities. In 1954, they began planning a new ticket facility for Pennsylvania Railroad Station in New York City, one of the largest railroad passenger terminals in the world, handling nearly a half-million passengers daily. This made necessary a system utilizing methods which were more efficient than those previously used in Pittsburgh or Philadelphia. After considerable study of the performance and cost of several systems, closed circuit TV (television) was decided on to handle the storage and distribution of the vast amount of information necessary for the sale and reservation of space on Pennsylvania Railroad trains.

System Description

Referring to Fig. 1, the ticket facility has three areas between which TV is used to transfer information: the ticket counter where direct sale of tickets for reserved space is handled, the telephone positions where reservations are taken by telephone, and the coupon file center where coupons representing all reserved space are held until requested by ticket agents or the telephone clerks.

The system performs three main functions. First, it supplies information on available reserved space to all operating positions in the facility. Since the information is readily available, this eliminates any possibility of wasted effort requesting space that has already been sold, and it is of basic importance in reducing the work load on a ticket reservation system. This portion of the system is capable of having all positions select the same camera at the same time. Second, a 2-way audio and 1-way video system between the ticket agents and the coupon file center is used for agents to request space. Third, a system to flash coupons between the coupon file center and the telephone reservation operator is used to confirm space requested by the operators. When the system is not used for these functions, it is used to provide other information such as train arrivals, orders of the day, and institutional advertising on the ticket counter monitors.

The telephone reservation operation is shown in Fig. 2. The customer *A* calls in to reservation clerk *B*, who dials the proper space availability camera *C* to check and determine if general space is available for the date and time requested by the customer. The clerk *B* then dials the coupon flash-back center *D* and requests the customer's space over the TelAutograph system. The clerk at *D* pulls the proper coupon, which represents a specific space and places it in front of the coupon flash-back camera. This camera flashes the coupon back to a monitor located at the reservation clerk's position *B*, who then checks to see if the specific space given falls within the limits of the general space requested. Clerk *B* then tells the customer what specific space has been given to him, i.e., roomette number and car number. This coupon representing the specific space is filed in the coupon file until the customer either purchases or cancels the space. For the customer to pick up his reserved space he must come to the ticket counter where the ticket agent will request the space. The procedure is the same for the direct sale of tickets described in the following paragraph, except that dialing space availability is not required.

For direct sale of tickets the operation follows Fig. 3. The customer *A* comes to the ticket agent *B* and requests space on a train for a certain date. The agent *B* dials space availability system *C* to check if the space is available. He then dials the coupon center via the audio-video system. The audio is used by the agent *B* to request the customer's space. The video permits the customer to see the space being obtained by the coupon center clerk *D*. This space is then sent to the agent via Western Union Intrafax.

While this method of processing the ticket may seem complicated, it must be realized that the earlier method required more steps without the availability of rapid communications. A major contribution of the audio-video system is the psychology applied to the customer who sees action taking place on the processing of his ticket. The dynamics of this operation has reduced the impatient complaints to practically zero.

System Requirements

Information gathered from the previous system in the Pennsylvania Station and the new systems at Pittsburgh and Philadelphia was used to determine the number of operating positions to handle sale and reservation of space efficiently. Seventy-two operating positions were decided necessary to handle telephone requests for reservations. Each position is equipped with a video monitor, TelAutograph transmitter and dial selection of the video and TelAutograph circuits.

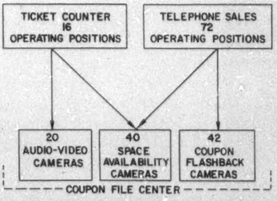


Fig. 1. Block diagram of system

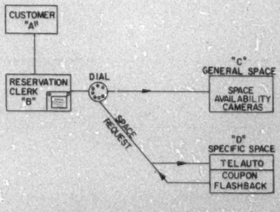


Fig. 2. Telephone reservation operation

Paper 60-260, recommended by the AIEE Broadcasting Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., January 31–February 5, 1960. Manuscript submitted November 25, 1957; made available for printing December 14, 1959.

J. W. ALINSKY is with the Dage Television Division of Thompson-Ramo-Wooldridge Inc., Michigan City, Ind.

The author wishes to express his appreciation for a large portion of the details of this paper to William J. Toeller, who was resident field engineer on the Pennsylvania Station project for the first year of operation.

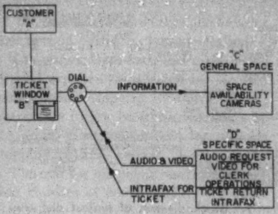


Fig. 3. Ticket pickup operation

For direct sale of reserved space, 16 positions are required, each equipped with a microphone and speaker, a monitor and selector equipment located at the ticket counter. Four additional miscellaneous positions selecting only space availability were used in various locations. Also, two positions, designated "Special Handling," are necessary for persons requiring special service where transactions could be handled either by direct sale or space reservation.

The number of cameras necessary for each system, i.e., space availability, coupon flash-back, and audio-video, were determined previously by the number and sizes of trains involved and the statistical data available for past experience on the popularity of each schedule.

The space availability system is most critical, requiring the reproduction of symbols and representing reserved space for periods to 8 weeks on one schedule of trains and to 16 weeks on another schedule. Results of experimental work showed that this information could be handled on 36 specially designed charts, each covering a period of 2 weeks. To provide for future expansion, 40 camera positions were designed into this system.

Forty-two cameras were used for the coupon flash-back system and 20 cameras for the audio-video system. These cameras are selected within each group through a hunting system, whereby if the first unit in a group is busy, the second is selected, etc. If all are busy, redialing is necessary.

System Engineering

To work out engineering details, the system was divided into the following parts: camera, video distribution, sync generators and sync distribution, power for distribution system, monitors, audio system, and switching.

CAMERA

Specifications for the camera were: not less than 400-line horizontal resolution, 0.5-volt peak-to-peak noncomposite video output, driven horizontal and vertical sweep for maximum stability, self-contained power supply, and all controls located on the camera. Video test points should be included which are available from the exterior of the camera. Roof prisms are used on the cameras to provide their better positioning and also to allow a horizontal placement of the camera to minimize vidicon spot damage.

VIDEO DISTRIBUTION

A video distribution amplifier had to be designed to meet the requirements of video distribution for three separate groups of cameras. In the space availability system, a camera would have to feed 99 monitors (which would provide a system expansion of five additional monitors). This assumes that all positions might require information from the same camera at the same time. In the coupon flash-back system, one camera is capable of feeding only one monitor at a time out of the 74 monitors using this system. Here the distribution amplifier must supply isolation between switching units. The audio-video system is similar to that of the coupon flash back except that video distribution is provided to only 18 monitors. In addition to these electrical specifications, the distribution amplifier had to be designed so that it was easy to maintain.

To solve these problems, it was necessary to have an amplifier supplying 99 outputs at an impedance of 75 ohms with sufficient isolation between outputs to prevent loss of high-frequency components in connection with line capacity, and to minimize any effect caused by line reflection.

Fig. 4 illustrates the output circuit of the distribution amplifier capable of feeding 25 75-ohm lines at 0.5 volt peak to peak. The load on the amplifier is 6 ohms and consists of the 25 lines, each separately connected to the amplifier through a 75-ohm isolation resistor. This method provides about 22 decibels of isolation between lines, which was sufficient to provide a clean picture.

A distribution amplifier constructed by combining four of these channels on a module or plug-in chassis was capable of meeting all specifications. By bridging the inputs of the channels together, 100 outputs were made available from one input to meet space availability camera requirements. By dividing the 74 positions which used the coupon flash-back system into groups of three, one channel could supply video to the resulting 25 groups. For the audio-video system, one channel could directly feed the 18 positions.

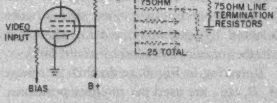


Fig. 4. Output circuit of video distribution amplifier

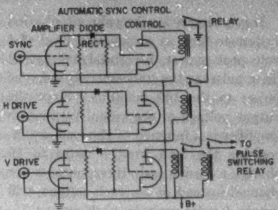


Fig. 5. Automatic sync control

Fifty-six distribution amplifiers were necessary for the video distribution system. Constructing the distribution amplifier as a module helped solve the maintenance problem, since a defective unit can be replaced by plugging in a spare unit.

SYNC GENERATORS AND SYNC DISTRIBUTION

For the synchronizing system, two Dage *410-A* sync generators are used supplying Radio-Electronics-Television Manufacturers Association pulses to the equipment. These two units are controlled by an automatic changeover which provides protection to the equipment against pulse failure.

Referring to Fig. 5, the automatic changeover operates by sampling three pulses from the generator in service, the vertical drive, the horizontal drive, and the sync, each feeding separate channels. The pulse is amplified in half of a twin triode and rectified by a diode. The resultant voltage is used as bias on the second half of the triode which operates as a control tube with a relay in its plate circuit. With pulses supplied, three relays are energized, closing single-pole single-throw contacts. The contacts of the relays in the three channels are in series with the coil of a control relay which energizes a solenoid-operated rotary switch selecting the proper sync generator. The automatic control unit is designed to switch generators when any component in the unit or the pulse which is feeding it fails.

It is necessary for the pulse distribution system to feed, at a minimum level of 4 volts peak to peak, horizontal and vertical drive pulses to the 100 cameras, and sync pulses to the 100 monitors. Since the output voltage of the video distribution amplifier is a function of the load, by increasing the load from 6 ohms to 50 ohms, i.e., using only three outputs per channel instead of 25, each output will supply 4 volts peak to peak into a 75-ohm line. With cameras divided into 12 groups and the monitors into 12 groups, and all units in each group bridging off the line, two distribution amplifiers are necessary for pulse distribution to the cameras, and one for sync distribution to the monitors. Using this method of synchronizing instead of composite video for locking in the monitors to cameras provides optimum performance during switching and eliminates any vertical bouncing or horizontal tearing.

Power for Distribution System

Since self-contained power supplies on each distribution amplifier would have made a unit too large or bulky for plug-in operation and also would have doubled the amount of rack space necessary for the system, the filament transformer was the only power equipment mounted on each chassis. Plate power for the 59 video and pulse distribution amplifiers totaled 30 amperes at 150-volts d-c and was supplied by two 40-ampere 150-volt d-c copper oxide rectifiers. These two rectifiers are controlled by a timer changeover unit which switches rectifiers once daily. Also included is a relay, controlled by the output voltage of the rectifier, which switches rectifiers in case of failure. Filtering for the rectifiers consists of a capacitor input filter with 1,000-μf (microfarad) 200-volt input condenser and a 5,900-μf 200-volt output capacity which is made up by 100 μf in each of the 59 distribution amplifiers. The 100-μf capacitors serve a dual purpose in this application. In addition to serving as filters for the d-c supply, they also provide necessary isolation and decoupling of the d-c line between distribution amplifiers.

Monitors: A 14-inch monitor is used at the operating positions, utilizing a video amplifier of an 8-megacycle bandwidth and a gain of at least 45 decibels. This provides a high-resolution high-contrast picture with a noncomposite video input of between 0.3 and 0.5 volt peak to peak. The monitor must be capable of a high-luminance raster while maintaining good over-all focus. For maximum stability during switching, the monitor sync circuits have to be driven. A Dage *602* monitor is used, which provides sufficient video gain and bandwidth to meet these specifications, plus a 12-kv regulated high-voltage power supply to prevent blooming at high brightness levels.

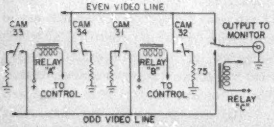


Fig. 6. Space availability video relay panel

Audio System

The audio system which supplies the 2-way communication link between the ticket counter and the coupon file center should provide sufficient gain to permit operation within 3 feet of microphone or speaker. The amplifier used is designed for low-impedance input, and an output of 500 ohms, since long lines between microphone, amplifier, and speaker are used. The input stage is built around a *GE 2N78* n-p-n junction transistor in a grounded emitter circuit which provides an input impedance around 500 ohms, thus eliminating the need for microphone input transformer. The remaining stages in the amplifier are resistance-capacitance coupled triodes.

Switching

The switching system is divided into two parts: 1. video switching for the space availability cameras, and 2. selector units for each monitor position, which control the availability video relays and switching circuits for either the coupon flash-back or audio-video systems.

Space Availability Switches

The space availability system requires switching video from 40 distribution amplifiers to 99 monitor positions. This system had to be designed so that any number of these positions could select the same camera simultaneously while maintaining a 75-ohm termination impedance.

Instead of one relay used for each video signal to be switched, a system was devised whereby two video signals were selected by one relay, cutting the total number of relays needed by about 50%.

Referring to Fig. 6, to do this, 21 relays *A*, *B*, etc., are used per monitor positions, each switching two different video signals from the distribution system to two separate lines (odd line and even line), which are common to the 21 relays. On one of these lines the odd-numbered cameras are switched, i.e., 37, 39, 41, etc., and the other line receives the even-numbered cameras, i.e., 38, 40, 42, etc. The video from each of these lines is then selected by an odd-even relay *C*, and fed to the monitor. The selection of a camera requires two operations: the energizing of the proper video relay, and either energizing or nonenergizing the line selector relay. The method of selection will be explained in the next section.

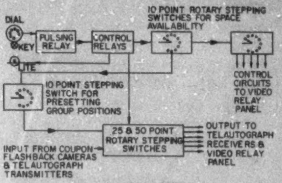

DIAL
KEY
PULSING RELAY
CONTROL RELAYS
10 POINT ROTARY STEPPING SWITCHES FOR SPACE AVAILABILITY
LITE
10 POINT STEPPING SWITCH FOR PRESETTING GROUP POSITIONS
CONTROL CIRCUITS TO VIDEO RELAY PANEL
25 & 50 POINT ROTARY STEPPING SWITCHES
OUTPUT TO TELAUTOGRAPH RECEIVERS & VIDEO RELAY PANEL
INPUT FROM COUPON FLASHBACK CAMERAS & TELAUTOGRAPH TRANSMITTERS

Fig. 7. Block diagram of typical dial controlled selector unit

Selector Unit

A method of selecting circuits for the coupon flash-back and audio-video systems and some means of controlling the availability video relays were necessary. This method had to be simple, since unskilled workers were to operate the equipment.

Push-button selection, which has been predominantly used in the industry, was discarded because of the size and wiring complexity where a large number of circuits were to be controlled. A dial pulse controlled system similar to a telephone system is employed instead, providing a method which is familiar to everyone.

At each operating position, a dial, an indicator light, and a dial release key are mounted, performing the same function as the dial, tone, and cradle of a telephone system.

Referring to Fig. 7, by turning the dial key to dial position, the pulsing and control relays are energized, lighting the indicator light. Dialing a 2-digit code operates the control relays in the proper sequence actuating 10-, 25-, and 50-point rotary stepping switches.

Space availability video relays are controlled by two of the 10-point stepping switches. One operates as a function of the ten's digit, the other a function of the unit's digit. When dialing an odd number, i.e., 37, 39, 41, etc., both stepping switches operate to select the proper video relay. The odd-even relay is energized by the unit's stepping switch. When dialing an even number, i.e., 38, 40, 42, etc., a similar action takes place except that the odd-even relay is not energized.

The coupon flash back and audio-video circuits are directly selected by 25-point and 50-point stepping switches. These operate as rotary hunting switches with the group positions marked by 10-point stepping switches.

The unit for telephone reservation position consists of control relays for space availability switching and stepping switches for coupon flash-back switching.

The ticket counter switching unit combines control relays for space availability and stepping switches for the audio-video system.

Some Pertinent Statistics

Over 60 miles of wire and coaxial cable were used in the wiring of the racks. The system employs 3,000 tubes and more than 3,500 relays and switches; more than 100,000 connections were used in the installation of the system. Reliability of the system has been more than gratifying; it has been in operation for approximately one year, representing over 100 camera years and 100 monitor years of operation. The total repair time on the cameras was less than 33 hours; the total repair time on the monitors was approximately 110 hours. The maintenance time on the cameras, which includes checking of tubes, cleaning of lenses, dusting and adjustment of controls, was less than 500 hours for the year.

One of the greatest areas of service problems was in the switch modules where over 500 hours were used in cleaning switch contacts and repair. Most of this is attributed to the fact that the equipment was installed concurrently with a rebuilding construction program within the station allowing much loose plaster dust, sawdust, etc., to accumulate within the racks. It is interesting to note that all the other equipment, which includes power supplies, sync generators, distribution amplifiers, etc., had a total repair time of less than 20 hours for the year.

A few major problems were encountered during the first few weeks of operation, the most noticeable of which was crosstalk in the video of the space availability switching system. Examination revealed that the video current flowing through the ground busses on the video switching panels was developing an excessive amount of voltage across the busses. Installing 1/8- by 1-inch ground straps across the existing busses bypassed this current to the main ground and eliminated the crosstalk.

A Military 4-Wire Field Telephone Incorporating an Electronic Signalling System

J. L. FAHERTY
NONMEMBER AIEE

A. S. HOWELL
MEMBER AIEE

A NEW field telephone has been designed for use with the 4-wire electronic switchboards now being developed for the U.S. Army Signal Corps. This local battery instrument incorporates new features in systems applications. All telephone signalling to and from the switchboard is accomplished on a voice frequency basis, thereby providing compatibility with radio and wire carrier. The digit tone signalling plan is readily adaptable to the use of a key-type selector set. The latter affords more rapid digit setup speeds than are possible with the use of the conventional rotary dial. The introduction of a 4-page wire system eliminates critical hybriding in the telephone as well as throughout the switching centers.

Compatible functioning of this telephone with the above-mentioned electronic switchboards requires locally powered compound tone generators for outgoing signalling and replacement of the conventional bell ringer by electronic means to produce a station-audible signal on incoming calls. These functions are accomplished through the application of solid-state devices.

Several other features have been incorporated including a provision for extension telephones and built-in facilities to convert for point-to-point operation. In addition, improved voice transmission is obtained from the replacement of the conventional carbon microphone with a high-quality magnetic transducer.

The pattern of this instrument reflects the new point of view of the military laboratories in field set packaging. A carrying case of low, flat silhouette containing and supporting all apparatus is designed to eliminate the need for a separate carrying bag. A new lightweight molded handset is secured to the supporting cradle by magnetic means. Consistent with modern circuit-packaging techniques, all circuit assemblies are of the plug-in type, making maximum usage of printed wiring.

Service test production units of this new 4-wire electronic telephone are now being completed by the Stromberg-Carlson Company. These instruments, together with the newly developed electronic local, long-distance, and tandem switchboards will be field tested at Fort Huachuca, Ariz., by the U. S. Army, during 1960. This paper will describe the engineering considerations and some of the problems involved in developing an instrument to meet the broad objectives outlined above.

Application

The primary function of the 4-wire field telephone is to provide reliable voice frequency signalling and voice transmission between the user and the local electronic switchboard, when interconnection is by means of nonloaded 4-wire field cable. The operating range is up to a 25-db (decibel) maximum loop loss at the highest signalling frequency assuming cable attenuation characteristics similar to those of *WDI-TT* field wire. Digital signalling to the switchboard is

---

Paper 59-259, recommended by the AIEE Wire Communication Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 1-6, 1959. Manuscript submitted November 28, 1958; made available for printing September 18, 1959.

J. L. FAHERTY is with the U.S. Army Signal Research and Development Laboratories, Fort Monmouth, N. J., and A. S. HOWELL is with the Stromberg-Carlson Division of General Dynamics Corporation, Rochester, N. Y.

The solutions for problems occasioned by the nature of this project represent the combined thinking and advice of many engineers too numerous to mention. The writers gratefully acknowledge this assistance from members of the Telecommunications Division of the Stromberg-Carlson Company, and the U.S. Army Signal Research and Development Laboratories.

CODE	T	U	V	W	X	Y	Z
FREQUENCY (CPS)	1700	1900	2100	2300	2500	2700	2900

DIGIT	CODE	FREQUENCY (CPS)
1	VW	2100 / 2300
2	WX	2300 / 2500
3	UY	1900 / 2700
4	UV	1900 / 2100
5	XY	2500 / 2700
6	WY	2300 / 2700
7	VX	2100 / 2500
8	UW	1900 / 2300
9	VY	2100 / 2700
0	UX	1900 / 2500
R (RECALL)	WZ	2300 / 2900
SEIZE	T	1700
RELEASE	YZ	2700 / 2900

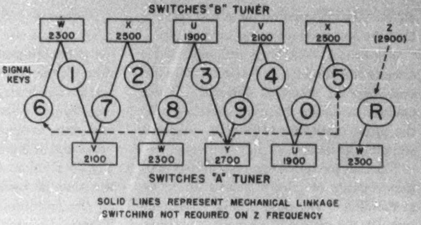


Fig. 1 (left). Signal frequency assignments

Fig. 2 (above). Switching plan for 5×5 keysender

accomplished by compound tones using a two-out-of-five frequency code. Preference was given to a system whereby the compound tones are applied simultaneously rather than sequentially. Such choice was favored because of the attendant reduction of mechanical complexity in the keysender and minimization of the effects of fading when a radio link is interposed between the local and long-distance switchboards.

To accommodate all switchboard signalling functions, including the digits, seven frequencies extending from 1,700 to 2,900 cps (cycles per second) are utilized. These frequencies are spaced 200 cps apart as a result of practical filter design considerations. The requirements for teletype transmission in the frequency range of from 850 to 1,650 cps dictated that this allocation be chosen for the signal tones. Otherwise, placement at the lower end of the speech frequency band would prove desirable because of the increase in nonloaded cable attenuation with frequency. The signalling plan may be illustrated by describing the operation of the telephone from the user's station.

Outgoing Calls

When the handset is removed from the cradle, a timed-out burst of seize tone is transmitted from the telephone. The receipt of this signal at the switchboard is acknowledged by the return of a 600-cps dial tone which occurs at the telephone receiver.

Upon hearing dial tone, the calling party may transmit the desired number by manual operation of the keysender. Each digit key selects a specific compound 2-tone signal. Direct operator access before digiting is obtained by depressing the *O* button after receipt of dial tone. Operator access during a conservation is obtained by means of a recall key and another compound frequency assignment.

At the conclusion of the conversation, the handset is replaced upon the cradle, thereby returning the hookswitch contacts to normal. At the same time, a compound release tone is applied to the transmitting line pair through action of a release signal control. This release signal will persist for several seconds, after which the telephone automatically returns to stand-by condition. When the telephone is in stand-by or on-hook condition only the tone ringer detection and the release control circuitry are active, drawing a low idling current in the microampere range.

Incoming Calls

An interrupted 600-cps ringing signal generated in the switchboard is amplified and detected in the telephone. The detector activates a 2,700-cps generator and an amplifier through which the tone signal is applied to the handset receiver and coupled to an acoustical radiator to produce the station-audible signal.

When the handset is lifted from the cradle, seize tone again is sent out. The reception of seize tone at the switchboard removes the ringing signal. The latch-up circuit in the telephone operates under control of the seize signal to place the circuit in talking conditions.

During voice transmission, incoming signals on the receiving line enter a receiving amplifier provided with a manual gain control, then they are applied to the telephone receiver. Voice frequency currents from the microphone are amplified, then applied to the transmitting line pair. A unidirectional sidetone path is included between the send amplifier output and the telephone receiver.

Signal Frequency Assignments

In the signal frequency assignments, shown in Fig. 1, the make-up of the first two digits is chosen to include a common frequency. This simplifies the register used in routing calls to local lines or trunks for prefix numbering of digit 1 or digit 2. Otherwise, the specific digit frequency assignments are arranged to simplify the keysender design, as will be described. An additional system requirement is that the release compound not contain the seize frequency. The latter provision prevents reseizing of the switchboard when disconnect occurs before the decay of the release signal.

Keysender Switching Plan

The keysender provides the means to set up the digit and recall signals by suitable switching of frequency-controlling tank circuit elements in the two telephone signal generators, designated *A* and *B*. To reduce the number of switches and thereby effect economies of space and cost, mechanical compounding is employed. The switching plan, illustrated in Fig. 2, is designated to permit the operation of a particular tuner switch by adjacent keys using a common mechanical linkage. Thus only 10 tuner switches are associated with the digit keys, where otherwise 20 switches would be necessary.

Several further considerations are involved in the keysender plan. It is preferable that the tuner switching described be accomplished on a "dry" basis, whereby the connections are established before the oscillators are powered. This is done through control

Vol.
Pt. 1
196

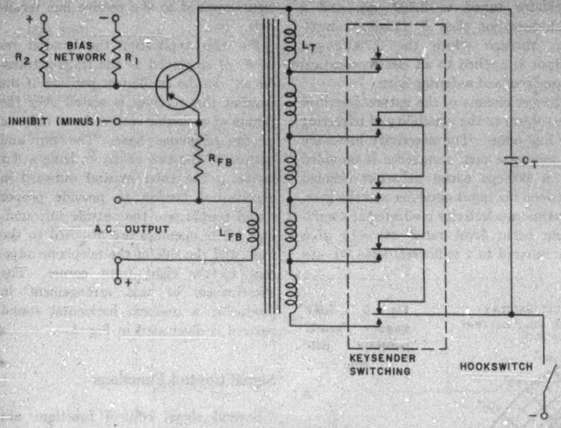


Fig. 3. Signal oscillator circuit

of the battery switch by a sequentially operated mechanical link common to all signalling keys. To obviate the need for a separate seize tone generator it is necessary that one of the signal oscillators be tuned to the seize frequency when the keys are in the nonoperated position. Hence, the *Y* switch in Fig. 2 is used as a transfer switch operative from a common key link to shift the *A* tuning from the seize frequency to the basic tank frequency upon the operation of any key.

Human engineering studies were made to determine an optimum arrangement of the digit keys based upon operator performance. Subjective tests were conducted on three plans comprising placement of the buttons in two horizontal rows of 5, two vertical rows of 5, and four horizontal rows of 3, 3, 3, and 1. Although the findings from these tests indicated no difference among the key sets in the time required to send a telephone number, a significant reduction in the total number of errors was made with the use of the two horizontal rows-of-5 pattern. It was noted further that the digit selection speed for numbers between three and ten digits in length averaged two digits per second.

Signal Oscillator Design

Effective and reliable functioning of the signal oscillators, as well as all telephone circuitry, must be maintained over an operating ambient temperature range of −40 to +149 degrees Fahrenheit. System requirements dictated a signal frequency stability of ±20 cps and an essentially distortion-free output which can be controlled by mechanical as well as electronic switching means. In a portable, local-powered instrument it is essential that these functions be accomplished at maximum economy of current drain, components, space, and weight.

The basic oscillator configuration, shown in Fig. 3, is designed to utilize a general-purpose-type p-n-p germanium transistor having a beta value greater than 60 at 25 degrees centigrade. In this arrangement a frequency controlling resonant circuit is formed by capacitor *CT*, and an inductance *LT*, which is wound as a number of separate sections on a toroidal core. A feedback winding, *LFB*, part of the same magnetic structure, is placed in the emitter circuit where it serves the dual function of supplying the feedback to the amplifier and the output voltage to the load. Degeneration to limit distortion products is provided by resistor *RFB*. Mechanical switching control is obtained through the keysender and hook switch functions, whereas electronic control is obtained from a negative inhibit voltage which reverse biases the emitter circuit. The operating bias network is formed by resistors R_1 and R_2 in conjunction with the nominal 9-volt battery supply.

Much consideration was given to the process for frequency-tuning the oscillator tank circuit during manufacture. The desired frequencies could be set up either by padding out the capacitive element with the use of shunt trimmer units or by the adjustment of individual taps on the inductive element. The latter method was chosen because it yields important space savings through the use of fewer components, with over-all assembly costs no greater.

Tone Ringer

The design function of the tone ringer is to provide an adequate station-audible signal on incoming calls at minimum operational and stand-by current drains. For maximum audibility under battle conditions it is desirable that the output signal be of high harmonic content with maximum emphasis on frequencies between 2,000 and 3,000 cps. The radiated signal at 3 feet from the source must meet a minimum level requirement of 60 db above reference sound pressure exclusive of all components below 2 kc in frequency. In addition, the full-volume signal output must be audible 100 feet from the telephone in all directions in an ambient noise level of +60 db. A manually adjustable volume control must be provided which is capable of reducing the acoustic output to a barely audible level close to the telephone under quiet conditions.

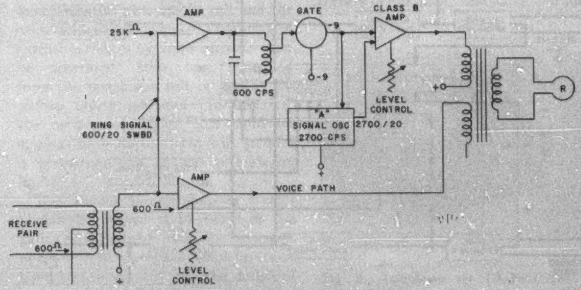


Fig. 4. Block diagram of tone ringer circuitry

The developed tone ringer circuit is designed for a high input impedance level to permit direct bridging across the receiving voice circuit. This eliminates the need for the switched removal of the ringer input when the telephone is off-hook. In the arrangement illustrated in Fig. 4, a 600-cps–20-cps interrupted ring control signal generated and cycled at the switchboard is applied at the ringer circuit input. The output of the amplifier is applied to a transistor gate. The latter is operative in the presence of ringing signal to turn on the *A* signal oscillator tuned to 2,700 cps, and a high-distortion class *B* transistor amplifier, through which the 2,700/20-cps output is applied to an electroacoustical transducer and radiating horn.

Ringer circuits of the nature described are subject to the possibility of triggering by line noise. The selectivity necessary to minimize such occurrence is provided by a 600-cps tuned network included between the input amplifier and the gate. Maximum selectivity is effected at a minimum input level rated at −15 dbm (db referred to 1 milliwatt) for a 600-cps signal applied to the receive line terminals.

For this application the handset receiver is employed as the ringing transducer. In the cradle rest position of the handset this receiver is seated over the throat of a double-section horn built in to the telephone base. The top and bottom of the two radiators have a flat section. The sides expand outward in exponential fashion to provide proper sound coupling to the outside surroundings. The openings are disposed to the front and the side of the telephone adjacent to the right front corner. The effectiveness of this arrangement in producing a uniform horizontal sound pattern is illustrated in Fig. 5.

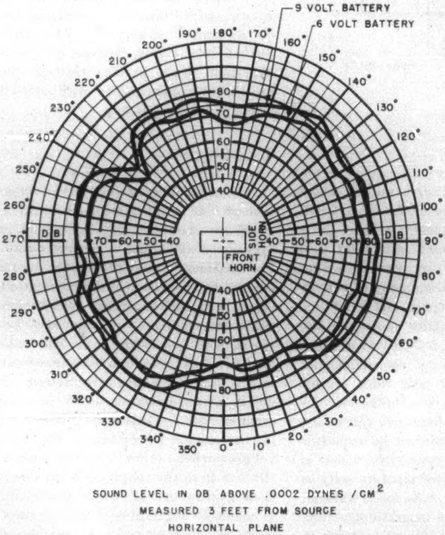


Fig. 5. Tone ringer sound radiation pattern

Signal Control Functions

Several signal control functions are necessary for the proper operation of this telephone with the switchboard. Of these, seize tone timing, latch-up, squelch, and release are automatic operations incorporated into the circuitry shown in the block diagram of Fig. 6. Such electronic means are employed in addition to the circuit controls provided by the hook switch and the keysender. These needs will now be described.

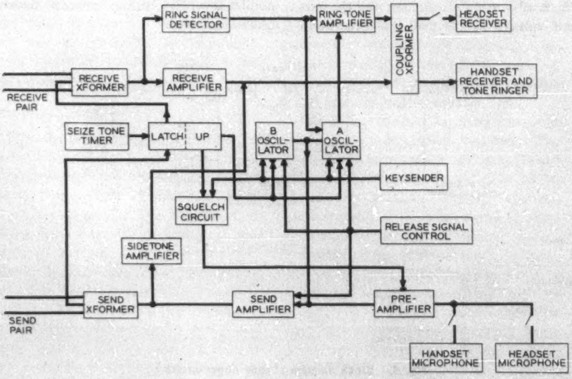


Fig. 6 (below). Block diagram of telephone circuitry

The application of timed-out seize signal from the telephone permits dial tone and busy signal to be returned for user recognition at a low level, nominally −18 dbm. The higher level of such supervisory signals, otherwise necessary to effect the electronic removal of seize signal at the telephone, would prove disturbing to the listener. The timing-out function is accomplished by means of a Zener diode regulated storage that controls a transistor gate. The latter gate drives a latch-up circuit of the 2-transistor type that operates to inhibit the signal generators.

The nominal voice output of the telephone is −9 dbm on a 1-kc tone basis for 28 dynes per centimeter² pressure applied to the microphone. The signalling output is −3 dbm maximum for compound tones. To prevent the simultaneous transmission of ambient noise picked up by the microphone, it is desirable to inhibit the transmitting preamplifier during outgoing signalling. On an incoming call similar inhibition is necessary when the handset is picked up during the receipt of 600/20-cps ringing tone from the switchboard. This is because the 1,700-cps seize tone, returned to the switchboard to trip the ringing, can be masked by ring signal entering the send path through the acoustical

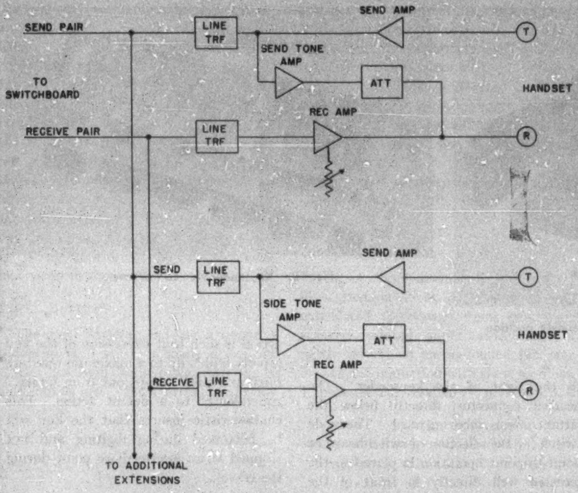


Fig. 7. Transmission path in extension service

coupling between receiver and microphone. These squelch functions are accomplished by a resistor-diode switch that is controlled by the latch-up circuit during seize signal transmission, and by the keysender when the digits are selected.

Following the completion of a call, the local switchboard link is disconnected by a compound tone sent from the telephone upon hang-up. It is desirable that this signal persist for several seconds to insure release in the presence of a radio fade. This release signal control is accomplished by means of a resistance-capacitance transistor gate. The capacitor in the timing circuit is alternately charged and discharged in the off-hook and on-hook conditions, utilizing hook switch contacts.

Hook switch control is utilized to apply battery to the operating circuits, to transfer the *A* oscillator for release or seize tone signalling and to charge and discharge the release control capacitor as described. Two additional disconnects, one between the ring tone amplifier and the battery, the other between the ring tone amplifier and the receiver-coupling transformer, were found necessary in the off-hook condition. These inhibit the otherwise undesirable effects quiescent transistor current on the latch-up control and on distortion during voice reception. The noted facilities are provided by four mechanically operated individual switches.

Extension Service

Several telephones may be bridged on the local line to provide extension service. The voice transmission pattern in such application is indicated in Fig. 7. It is noted that the unidirectional side-tone circuit included in the telephone is placed between the output of the send amplifier and the telephone receiver. Hence, the side-tone level is independent of the receive amplifier gain setting. Transmission between telephones on the extension line takes place at side-tone level in this 4-wire system. This level is nominally 9 db below the voice output appearing at the send line terminals.

Extension operation involves several necessary control functions between the first station to pick up the call and the other bridged telephones. Triggering of extension ringers by voice currents must be prevented. Seize tone generated from the telephones will be audible on pickup unless inhibited. In addition, the sending of release tone until all stations have disconnected will produce a premature trip of the switchboard link.

The control of extensions is accomplished by means of a simplexed d-c gating signal which inhibits the oscillators at the off-hook stations. These functions have been integrated into the design of the latch-up circuit discussed in conjunction with switchboard signalling. In this system the off-hook telephone sends out only trigger control power resulting in a nominal current of but one milliampere per extension.

Point-to-Point Operation

Point-to-point operation, as understood here, is communication between telephones without intermediate switchboards. Facilities provided in this instrument accomplish this service employing either voice frequency or d-c signalling. Voice frequency signalling is necessary when a radio link is interposed between stations; however, the signalling range over cable facilities alone is limited to approximately 4 miles. For operation over long cable loops, up to 20 miles, d-c control of the ringing detector gate is applied to the line on a simplexed basis. Such means are similar to those described for extension control.

A mode switch includes three positions to select switchboard or either method of point-to-point operation. For point-to-point communication the recall key is used for signalling, while both seize and release control functions are made inoperative. During voice frequency signalling, the signal oscillator frequency is shifted to 600 cps, the tuned frequency of the ringing detector. Otherwise, the telephone circuits function as described for switchboard operation.

Instrument Packaging

The form of the new instrument in the packed condition is illustrated in Fig. 8. For ease of stacking, a rectangular shape was chosen and proportioned for maximum stability in the horizontal rest position shown. To the same end all projections beyond the exterior walls of the box have been eliminated. Thus, the posts for carrying strap attachments are located in cutouts in the two rear corners of the base. Similarly, the line-

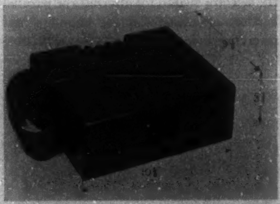


Fig. 8. Telephone set TA-341( )/PT in packed condition. Dimensions in inches: height $3^3/_4$, width $10^5/_8$, depth $9^1/_8$

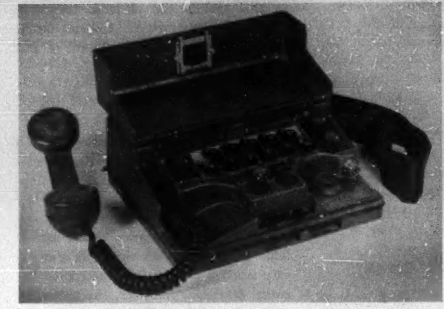


Fig. 9. Telephone set TA-341( )/PT in operating position

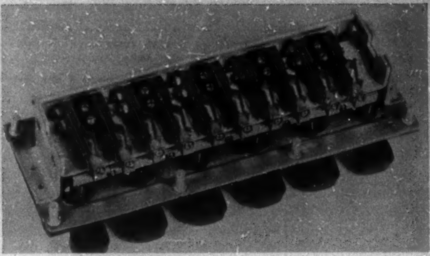


Fig. 10. Keysender viewed from rear

binding posts are placed in an *L* recess in the rear wall of the case. The booted latch provided in the front cover is flush with the wall surface.

The carrying case is of 3-piece construction comprising a cover, body, and base. These parts are molded of high-strength glass-filled polyester material. The base, including the battery well, is secured to the body in assembly upon addition of a waterproof gasket.

Opening of the hinged cover, by actuation of the one hand-operated latch, quickly changes the telephone from the packed to the unpacked condition; it is then ready for use. This operating position is shown in Fig. 9: here the handset has been removed from the cradle to expose the cushion seat and horn coupling for the handset receiver in the right foreground. The rectangular openings visible in the base expose the front section of the double horn radiator disposed to the front and the right-hand side of the box.

All panel controls and accessories are placed for maximum operating convenience in so far as feasible. Thus, the keysender, containing ten digit keys and a recall button, is centrally positioned. To the left of the keysender is located the selector switch for alternate or parallel headset operation. Ringer loudness and receiving voice level controls are visible to the right of the keysender. The headset connector, directly below the latter controls, faces outward. The mode switch for the selection of switchboard or point-to-point operation is placed in the recessed well directly in front of the keyset.

Major Functional Assemblies

KEYSENDER DESIGN

A rear view of the keysender is illustrated in Fig. 10 In this design all electrical switches are of the leaf spring type employing twin contacts. Spring pile-ups are cam-driven by sliders associated with each switch, and actuated from pivoted key toggles. The latter are individually waterproofed by rubber bonding at the pivot points. Crossbars on the toggle arms engage and move adjacent sliders to provide the mechanical compounding previously described.

Two important human engineering factors were considered in the areas of key spacing and key action. Thus the staggering of the keys on the upper and lower rows develops maximum inter-button spacing without increase in keyset height. Wide button spacing minimizes the danger of false digit selection when arctic glove liners are worn. Similar operating facility is designed into the key movement. The action is such that depression of the key rapidly builds up to a maximum resisting force and then "drops out from under," equivalently to a detent action. This characteristic insures that the key will be bottomed during digiting and not stopped at an intermediate point during the travel.

HANDSET DESIGN

The new lightweight handset is molded of a high-impact thermoplastic material. In addition to the transducers it includes a plate of magnetic material secured to the flat section of the handle. The latter provides magnetic handset hold-down by attraction to permanent magnets in the telephone case. The military *TA-235/PT* ring armature type of telephone receivers are utilized as transmitter and receiver elements. For microphone use a substantially linear frequency response characteristic is obtained by external low-frequency acoustical equalization.

MAGNETIC HOOK-SWITCH

The magnetic hook-switch illustrated in Fig. 11 represents an effective solution to a complicated design problem. Service conditions necessitate the location

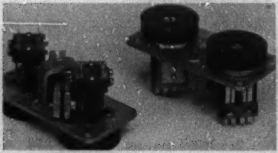


Fig. 11. Magnetic hook switch

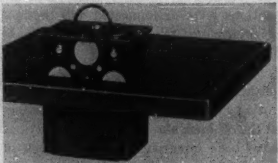


Fig. 12. Battery holder

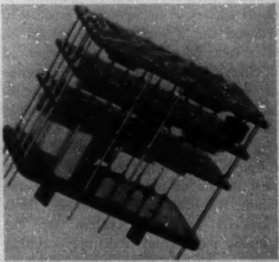


Fig. 13. Typical circuit module before encapsulation

of the switch on the inside of the case, and operation through a waterproof barrier. At the same time, magnetic hold-down of the handset is desired. In the latter case the flux change from handset removal may be utilized to rotate an armature. Applying this principle, the permanent magnets for handset retention are included in the switch, thereby eliminating any need for complex booting and mechanical linkages. The armature operates to actuate four switches of the leaf spring pile-up type. It is dynamically balanced to resist motion from shock.

Local Battery

The local battery, yielding a nominal potential of 9 volts, is made up of six *C*-size flashlight cells replaceable in a gasketed holder. Such arrangement facilitates the provision of battery taps. The use of three such taps has proved advantageous in accomplishing the circuit functions at minimum total current drain. The holder is inserted from the base of the telephone. Fig. 12 shows a bottom view of the molded base with the battery holder partly inserted into the waterproof well. As here applied, the expected capacity of this battery is 2,500 3-minute calls before replacement.

Circuit Packaging

Plug-in circuit modules are of the form illustrated in Fig. 13. These are encapsulated throwaway items employing stacked printed wafer assemblies. A newly developed spring-loaded pin jack is used in conjunction with these modules. Except for the oscillators, mounted in a separate package, the circuitry is divided into four units of similar size on a functional basis.

Conclusions

The new equipment described incorporates many new concepts in telephony. It is expected that the improved operating speed, reliability, and better transmission characteristics will prove of advantage to the military user.

References

1. Considerations Pertaining to the Design of an Electronic Telephone Switching System, **B. Brightman, M. P. Tubinis.** *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 78, 1959 (Jan. 1960 section), pp. 777–86.

2. Transmission Aspects of an Electronic Switchboard Employing Time Division Multiplexing, **J. C. Perkins, Jr.** *Ibid.*, pp. 949–53.

A Delay Distortion Simulation Set

FRANCES C. DUNBAR
NONMEMBER AIEE

Synopsis: The delay simulator is a test instrument which produces delay as a function of frequency in the voice frequency range: 300–3,300 cps (cycles per second). Its primary function is to simulate the delay distortion of Bell System transmission facilities, including open wire, loaded cable, and carrier systems. This is accomplished by combining four variable delay characteristics. The wide variety of delay characteristics thus obtainable may be useful for testing actual data transmission systems or any other waveforms.

THE DEVELOPMENT of data transmission systems for use over Bell System facilities requires techniques for the determination of effects of the transmission medium on system performance. It is hard to predict analytically the results of amplitude and delay distortion on complicated waveforms. For this reason, it has been customary to conduct tests with actual lines rented from operating companies. This method has some obvious drawbacks. Reproducibility of the transmission line depends upon switching through the same cable circuits and carrier channels. Even if this were readily accomplished, there would be variations in noise and temperature effects.

These disadvantages could be avoided by using electric networks to simulate the transmission characteristics of Bell System facilities. A scheme with sufficient flexibility would also be valuable as a standard for determining limits on distortion for an operable data system. The initial step in the implementation of such a plan was the design of a delay simulator This instrument produces delay as a function of frequency at voice frequencies—300–3,300 cps. Included within the scope of the simulator are open wire, loaded cable, and carrier systems. These facilities exhibit delay versus frequency characteristics which are roughly U-shaped, but not symmetrical. Minima of delay occur over a wide range of frequencies, from approximately 600–2,200 cps. The simulator reproduces these delay curves for any practical length or combination of transmission media. Fig. 1 illustrates several delay characteristics. *H-44* and *B-88* cable are widely used in local cable circuits. The two carrier channels represent extremes of delay distortion produced by upper and lower channels of long-haul carrier systems. In each case, the delay distortion was obtained by subtracting the minimum delay from the absolute delay.

Realization of such a variety of delay characteristics demands a scheme which is both flexible and easy to use. This is accomplished by combining the four variable delay characteristics shown in Fig. 2. These characteristics introduce delay distortion in four overlapping frequency bands. Each one adds no distortion outside its own range (within the 300–3,300-cps voice frequency limits). The first characteristic provides delay which decreases from a maximum at 300 cps to zero distortion at 800 cps

$$\text{Delay}=A\left[\frac{(800-f)}{(800-300)}\right]^3$$

where $f=300\text{–}800$ cps

The second decreases from 300–1,600 cps

$$\text{Delay}=B\left[\frac{(1600-f)}{(1600-300)}\right]^{3/2}$$

where $f=300\text{–}1{,}600$ cps

The third increases from no distortion at 1,200 cps to a maximum at 3,300 cps

$$\text{Delay}=C\left[\frac{f(-1200)}{(3300-1200)}\right]^{3/2}$$

where $f=1{,}200\text{–}3{,}300$ cps

The fourth increases from 2,200 to maximum distortion at 3,300 cps

$$\text{Delay}=D\left[\frac{(f-2200)}{(3300-2200)}\right]^2$$

where $f=2{,}200\text{–}3{,}300$ cps

Paper 60-45, recommended by the AIEE Wire Communication Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., January 31–February 5, 1960. Manuscript submitted November 5, 1959; made available for printing November 27, 1959.

Frances C. Dunbar is with the Bell Telephone Laboratories, Inc., Murray Hill, N. J.

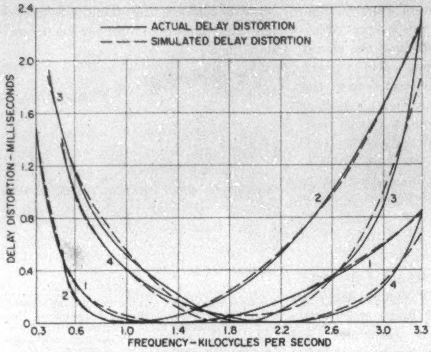


Fig. 1 Delay distortion of transmission facilities: Curve 1—100 miles of 19 gage cable, 0.062 microfarad per mile, 44 millihenrys loading, with two repeaters and composite sets. Curve 2—100 miles of 19 gage cable, 0.062 microfarad per mile, 88 millihenrys loading, with two repeaters. Curve 3—Maximum distortion of 500-mile J2 carrier system with one pair of terminals. Curve 4—One group connector, channel 1: common terminal equipment of K, L, and TD2 broadband carrier systems

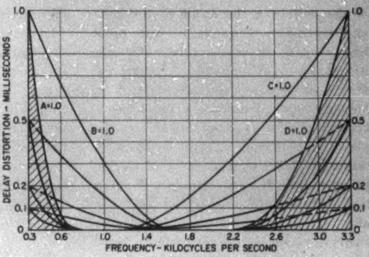


Fig. 2. Variable delay characteristics

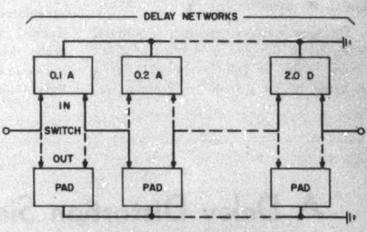


Fig. 3. Block diagram of delay simulator

The coefficients *A*, *B*, *C*, and *D* represent the maximum amounts of delay distortion added by each characteristic at 300, 300, 3,300, and 3,300 cps, respectively. Each coefficient has been furnished in units of 0.1, 0.2, 0.2, 0.5, 1.0, and 1.0 millisecond maximum delay. Therefore by combining these units, each coefficient may be varied in steps of 0.1 millisecond up to 3.0 milliseconds. Two additional units of 2.0 milliseconds (not shown on Fig. 2) have been added to the *B* and *D* series. The total maximum delay distortion which has been provided is 3.0 milliseconds each for *A* and *C* characteristics, and 7.0 milliseconds each for *B* and *D* characteristics.

Each coefficient (0.1 *A*, 0.2 *A*, etc.) is realized as a separate network which may be switched into the circuit independently of any other. When a network is connected into the circuit, it introduces delay which varies with frequency. In addition, it has an amount of delay which remains constant with frequency. This flat delay plus the delay distortion comprise the absolute delay of one coefficient. Each delay network also added some loss distortion as a result of dissipation in the component inductors and capacitors. It is apparent that any appreciable loss distortion could not be permitted in a device for studying delay effects. As a result, networks producing greater than ±0.2 decibel (db) distortion were equalized to some constant amount of loss. The deviation from flat loss for one coefficient ranges from ±0.02 to ±0.3 db. Equalization increased the loss of some of the networks to as much as 3.5 db. To avoid readjusting gain levels whenever a network is switched into the circuit, a pad with the appropriate loss has been inserted in the OUT position of each switch. The total flat loss of the simulator is 34.6 db, with a deviation which is determined by the coefficients in use. Therefore, for any combination of coefficients, the loss remains approximately constant and the absolute delay is the sum of the delay distortion and the flat delay of the coefficients. A block diagram of the simulation set is given in Fig. 3.

Some examples of simulated transmission lines are represented in Fig. 1 by dashed lines. The proper coefficients were determined as follows. An approximation was obtained in the 800–2,200-cps region by matching the required delay at 800–2,200 cps with the *B* and *C* coefficients, respectively. Where additional delay distortion was required at lower and higher frequencies, coefficients *A* and *D* were added. Sufficient amounts were selected to add the necessary delay at 500 and 3,000 cps. A wide variety of delay curves may be simulated in this way. For a narrower frequency range it may be expedient to readjust the coefficients slightly or to choose different matching

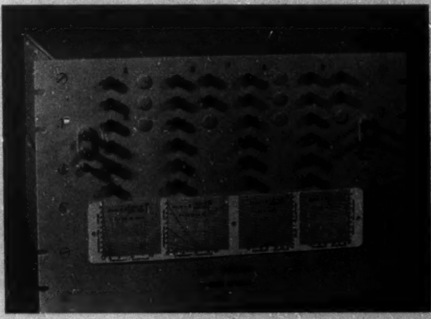


Fig. 4. Exterior view of delay simulator

Vol
Pt.
196

frequencies for one or more coefficients. If the coefficients for individual transmission lines are known, the delay distortion for any combination may be simulated merely by summing the coefficients. In the case of loaded cable, the delay is proportional to cable length (assuming typical circuit make-up), and therefore coefficients may be scaled from those obtained for one particular length. The delay of carrier systems depends mainly upon the number of carrier terminals and the channel allocations at these terminals. Consequently each system should be considered individually, or an average make-up may be assumed.

As previously mentioned, each coefficient is realized by a separate network. The network may contain from two to four all-pass sections and, in some cases, one loss equalizer section. These sections are all constant-resistance types which are unbalanced to ground. The impedance level is 1,000 ohms. All of the delay networks have been housed in special cans which are mounted on a front panel by means of supporting brackets. The pads are assembled on resistor mounting strips which are suspended beneath the top cover. Push buttons on the front panel actuate the switches that connect the delay networks or their associated pads into the circuit. Input and output terminals of the simulator are located on either side of the bank of push button switches. Ground terminals are insulated from the chassis but may be grounded to the front panel if so desired. Fig. 4 shows the simulator.

The value of this delay simulator lies in its flexibility and ease of operation. It provides delay characteristics which may be employed individually or in any combination. Furnishing delay distortion in four distinct but overlapping ranges has enabled the accurate simulation of a wide variety of Bell System transmission media. After the coefficients have been determined, any length or combination of facilities may be reproduced simply by scaling or addition of coefficients.

Redstone Arsenal Closed-Circuit Educational Television System

J. P. ARGO
NONMEMBER AIEE

I. N. HOWELL
ASSOCIATE MEMBER AIEE

THE REDSTONE ARSENAL at Huntsville, Alabama, is now pioneering in fields other than missile development. To illustrate this fact, the U. S. Army has had in operation for slightly over a year a unique and extensive educational TV (television) network. This network, with its associated administration, resembles very closely a nationwide broadcasting studio.

The Army desired a system for multipoint transmission from mobile TV studios with capabilities for programing on a video base-band basis and retransmitting on both a video and r-f (radiofrequency) basis. Furthermore, all outside plant had to be of aerial construction for security reasons. The r-f retransmission was to be confined to the U. S. Army base or post while the video retransmission was to be fed into the American Telephone and Telegraph Company's video facilities for delivery to Fort Knox, Kentucky, and other locations. In addition, the Army was providing another studio near the master-control room where the video was fed directly to the console. In an adjacent room, four Ampex Corporation video-tape recorders were also to be provided by the Army. The system uses components which have generally been found suitable for our company's purposes in providing educational TV service. While these system components have been adequate for normal educational TV requirements, where home-type receivers are used for viewing, no evaluation had been made where more rigid quality standards were desired. Fortunately, our company has had excellent co-operation from the Army and all others concerned with this system to assist in achieving the quality which is necessary in order that this type of educational TV system succeed.

System Description

This system as engineered consists of two separate networks on the Army base with the ability to program material received simultaneously over five separate video channels for purposes of retransmission over any one channel into the nationwide TV network on a closed-circuit basis. Of the two on-base closed-circuit systems, the one transmitting to the master control room is normally referred to as the transmitting network. This network, Fig. 1, consists of an aerial cable r-f amplifier layout which permits each mobile van equipped with high-quality image orthicon and studio vidicon camera chains to transmit on one of five channels from any of 21 transmitting locations over the network and into the master-control center. The channels in use are standard vhf (very-high frequency) 2, 4, and 6, plus channels 7 and 9 modulated downward in frequency to transmit over the cable system as subchannels between 36 and 52 mc (megacycles). There are several reasons for the use of subchannels. Five standard vhf channels in the channel frequency range of 2 through 13 require almost twice as much amplifying equipment. In addition, cable equalization problems are more complex. Also, the Army wanted the subchannels for a minor security technique in that those receivers not authorized for secret material were not to be converted from subchannels to standard vhf 7 and 9. This permitted programing the material on a selective basis for on-base consumption.

At the master-control center a group of extremely well qualified technicians, most of whom were formerly with major networks, strip the synchronizing pulses from the base-band video signals, reinsert new synchronizing pulses, and program the show for retransmittal to the second on-base system. Since the receiving system closely resembles the transmitting system, no drawing of it is included. The receiving r-f system receives information transmitted from the master-control room to some 29 receiving locations equipped with high-quality receivers. Simultane-

---

Paper 60-59, recommended by the AIEE Broadcasting Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., January 31–February 5, 1960. Manuscript submitted November 12, 1959; made available for printing December 14, 1959.

J. P. ARGO and I. N. HOWELL are with the Southern Bell Telephone and Telegraph Company, the former in Birmingham, Ala., and the latter in Knoxville, Tenn.

The authors acknowledge with appreciation the assistance given by C. A. Collins and A. D. Williams of Bell Telephone Laboratories, Inc., and Leon Davis of the American Telephone and Telegraph Company.

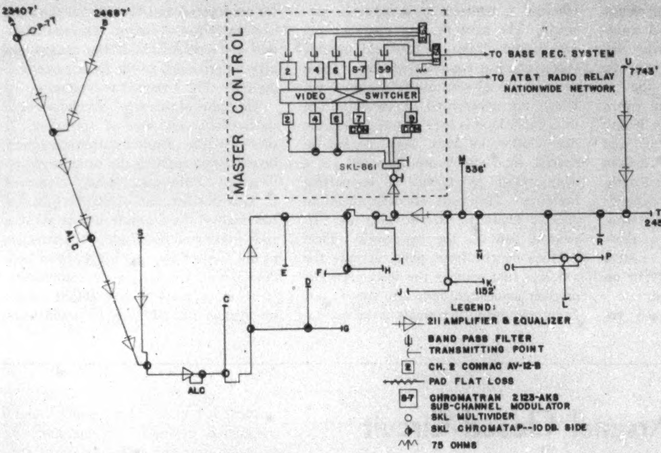

23407'
A
P-77
24687'
B
MASTER CONTROL
2
4
6
S-7
S-9
VIDEO
SWITCHER
2
4
6
7
9
SKL-861
TO BASE REC. SYSTEM
TO AT&T RADIO RELAY NATIONWIDE NETWORK
U
7743'
M
536'
T
2457'
ALC
ALC
S
C
D
E
F
G
H
I
J
K
1152'
L
N
O
P
R
LEGEND:
211 AMPLIFIER & EQUALIZER
BAND PASS FILTER
TRANSMITTING POINT
CH. 2 CONRAC AV-12-B
PAD FLAT LOSS
CHROMATRAN 2123-AKS SUB-CHANNEL MODULATOR
SKL MULTIVIDER
SKL CHROMATAP--10 DB. SIDE
75 OHMS

Fig. 1. Transmitting network

ously, one channel, selected by the director, is transmitted by radio relay to distant cities in the United States as required by the Army. On at least one occasion, the show was simultaneously transmitted to Fort Knox, the Pentagon, the Washington Press Club, and to New York, while also being received over the closed-circuit r-f receiving system on the base. At the distant cities a projection-type receiver is used and the video information is displayed on an 8-foot by 10-foot screen to permit large audience viewing for mass training purposes. Expansion of the video material for large-screen viewing requires that the transmitted signal be of excellent quality since the need for viewing very small electrical and mechanical parts is a necessity if training is to be as successful as if performed on the base.

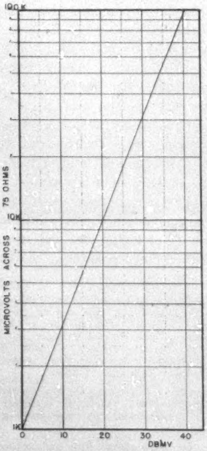


Fig. 2. Microvolts across 75 ohms versus dbmv

For the modulating source a Chromatron *2123-A* modulator was used to convert the video and audio signals received from the camera chain to r-f carriers. This modulator accepts a standard 1-volt peak-to-peak video signal from the camera chain and transmits it at nominally +40 dbmv (decibels referred to 1 millivolt); see Fig. 2 for dbmv. The output is filtered to reduce spurious frequencies and achieve better channel separation. One modulator is located in each mobile unit and terminated so that connection to the cable transmitting system can be made at any transmitting point without extensive adjustments. This connection is made through the use of a short length of triaxial *RG-11U*-type cable associated with the van. Triaxial cable is used to provide flexibility, to prevent radiation, and to improve reliability in the presence of radar. A short length of flexible cable is also secured to the utility pole and extends to a level of about 20 feet above the ground for connection to a Western Electric Company air dielectric 3/8-inch-diameter *CA-1878* coaxial cable. The *CA-1878* cable is supported by 6,000-pound tensile strength stranded wire to permit placing a ladder against it at midspan without danger of breaking. To minimize reflections no severe bends are permitted in the *CA-1878* cable.

Patch cords of the *RG-11U*-type triaxial cable are used to extend the *CA-1878* cable up and down the pole at each amplifier location or other locations where severe bending and flexing are necessary.

The amplifiers used for this system are

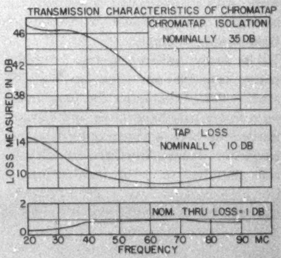


Fig. 3. Measured echo advantage of chromatap over multivider

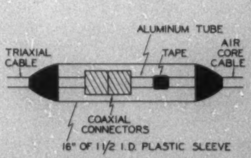
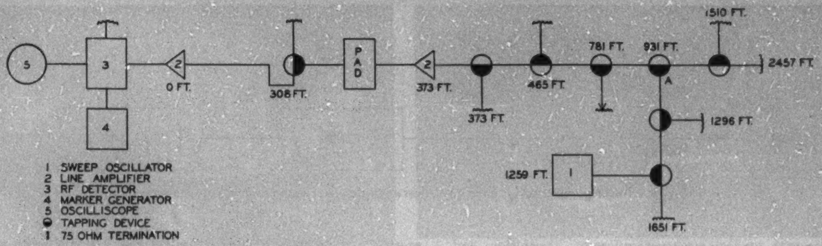
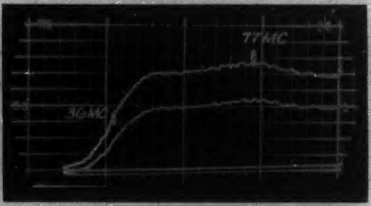

ALUMINUM TUBE
TRIAXIAL CABLE
TAPE
AIR CORE CABLE
COAXIAL CONNECTORS
16" OF 1 1/2 I.D. PLASTIC SLEEVE

Fig. 4. Sleeve used over coaxial connectors



(A)



(B)

Fig. 5. Cable section (A). Attenuation characteristics of cable section calibrated at 0.1-volt peak to peak, 2 db difference in levels (B)

of the *SKL-211* type and are mounted in metal cabinets at such a level as to permit maintenance from the ground. Associated equalizers, pads, etc., are also installed in these metal cabinets. Every third amplifier is equipped with an *SKL-830* automatic level control device to permit regulation of levels with temperature changes. No automatic slope equalizer has been incorporated in the system; therefore, seasonal adjustments are necessary for slope. The maximum level into the cable from any of the amplifiers is held to +42 dbmv. Levels higher than this tend to cause intermodulation. To keep noise within acceptable limits, the minimum input level to any amplifier was originally designed at +11 dbmv or better. Subsequent changes in the system, however, permitted reduction of this to +8 dbmv.

In the master-control room the various channels are demodulated through Conrac, Inc., *AV-12B* television tuners. From these the video is amplified by Western Electric Company *1-B* clamper amplifiers which also reduce low-frequency noise and distortion.

The Army is then given, at their master-control console, a standard 1-volt peak-to-peak video signal. The associated audio signals are separately cabled to the audio console. Throughout the portion of the system where the audio and video are utilizing the same facilities, audio levels are maintained approximately 10 db below the video levels. To facilitate level control a *77MC* pilot frequency is transmitted from one of the longer legs. The level of this pilot is maintained 5 db below the r-f carrier.

Other details of the design are in accordance with industry standards and have been covered in other papers.$^{1-3}$ It should be pointed out here that where band-pass filters are utilized they must be isolated from the cable network, preferably with an amplifier in each mobile unit, to avoid reflections which tend to degrade the echo rating of the system.

The requirements furnished by the Army were met with little difficulty with the exception of equalization at the sub-channel frequencies. The system was engineered assuming a flat 10-db loss for the tap side of the chromatap. Subsequently it was found that this item of plant, rather than having a flat loss, has the characteristics shown in Fig. 3. We are continuing to use the chromatap, however, because of its extremely good isolation characteristics. In the transmitting legs requiring several chromataps, with transmission required through the tap side, reverse equalizers were designed and constructed locally to achieve an over-all response characteristic flat within ±1 db from 30 to 90 mc. The channel characteristics for each video channel may, of course, be maintained considerably flatter than this. Another of the requirements of this system is that the signal from any transmitting point for any channel must arrive at the master-control room at essentially the same level. This requires extensive loss "building-out" and very close attention to maintenance to prevent unequal levels with associated cross modulation.

System Installation

Although the installation of a comparatively new type of r-f distribution system offered many unanticipated problems, previous experience by the Southern Bell Telephone and Telegraph Company in the placing and splicing of cables and the installation and interconnection of various types of equipment facilitated completion of this project on schedule.

Prior to placing the outside cables, discussions were held with supervisory and craft personnel responsible for the completion of the cable project. At this time the effect of such items as minimum bends, sheath breaks, moisture in cables, etc., was discussed. It was necessary to go even further with the cable splicers and give instruction on the splicing and on the terminating of solid and air dielectric coaxial cables.

Personnel designated to complete the installation and interconnection of the various items of equipment were selected on the basis of their ability in related work and the probability of using them in the line-up and maintenance of the system. Extremely close co-ordination was maintained between supervisory and craft employees during this phase of installation.

Redstone Arsenal, like many military installations, has continued to grow on an expedited basis. Consequently, its power distribution system and communications cables have been supplemented on a "when needed, where needed" scale. As a result, pole space available for the Southern Bell Telephone plant did not allow for maintenance of Bell System standards as to road and power clearances in a number of cases. An unusual amount of time was required by U. S. Army post signal and engineering groups, in co-op-

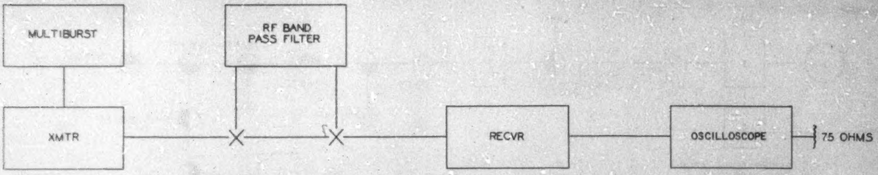


Fig. 6. Test setup for checking band-pass filters and video receivers in maintenance shop

eration with our own construction forces, to relocate and rearrange power fixtures and telephone cables to provide for standard clearances.

Since time was a most critical item throughout the entire project, the system was engineered on a crash basis and orders for equipment and outside cable placed immediately. In spite of these efforts, much overtime work was required to complete the installation prior to the required service data. As a result, line-up time was limited to the barest minimum.

In all cases, air dielectric coaxial cables were terminated in coaxial connectors on the messenger with triaxial cable used as patch cords to interconnect with tapping devices and amplifier equipment. Even though extreme caution was used in terminating all cables, a great amount of difficulty was experienced in securing water tightness at these points. This problem became quite serious during the cold rainy season but was finally solved by placing a short section of plastic tubing over these joints as shown in Fig. 4. The sleeve was installed in such a manner as to permit removal and reinstallation as required.

The placing of outside equipment cabinets at ground working level saved many man-hours of labor during the preservice line-up.

Because of the time element, it was necessary to form a co-ordinating committee consisting of a representative of the customer, the television equipment contractor, the television operating contractor, and the Bell Telephone Company. This group assumed the responsibility for authorizing deviations from original plans and on-the-spot solutions to problems as they arose. Without such co-operation from the interested parties, it is doubtful that the requested service date could have been met.

System Operation

The operation of this system can best be discussed under three separate categories: 1. preservice line-up, 2. daily operation, and 3. routine maintenance.

As expected, certain general problems were encountered common to the three areas of discussion, in addition to numerous problems which were unique to each particular category.

In the provision of a new service, technical in nature, an adequate supply of trained personnel is always a problem. In this particular case, time did not permit the establishment of formal training classes. Consequently, practically all training was done on the job.

Electronic circuit components and test instruments are mainly designed for use on TV community antenna systems. In some cases these provided neither the performance expected by personnel accustomed to broadcast-studio quality nor quality desired by the Army for this particular operation.

Ideally, this system should be essentially flat, between 36 mc and 88 mc, in order to pass the channels utilized. Little difficulty was experienced in obtaining flat response with ±1 db on channels 2, 4, and 6. This was not true, however, of the subchannel region. The design of the line amplifiers and tapping devices was such as to provide unwanted frequency distortion over the pass band. Because of the number of taps required in a limited geographical area, over-equalization resulted at a number of points. The cable layout shown in Fig. 5(A) was swept from 30–90 mc, the portion of the band now in use, in order to determine frequency attenuation characteristics, Fig. 5(B).

An attempt was made to utilize pure resistance tapping devices to improve frequency response at some of these points. It was possible to effect a definite improvement in this manner; however, it was necessary to add 75-ohm pads in the branches of these devices in order to provide isolation from impedance irregularities. The addition of pads necessitated the addition of line amplifiers to overcome increased losses.

This problem was resolved by a rearrangement of tapping devices which eliminated equalization effects on certain legs of the system. In a number of cases, as mentioned previously, it was necessary to design inverse equalizers with high-frequency losses in order to counteract over-equalization.

Two items of particular value throughout the preservice line-up and daily operation of this system, were a van-body truck equipped as a mobile shop and a Polaroid oscilloscope-adapted camera. The truck was equipped with work benches, storage lockers, power cords, a-c outlets, and r-f patch cords. Use of this truck enabled equalization, line-up, and maintenance work to continue regardless of weather conditions.

The Polaroid camera was used extensively to record sweep patterns during equalization and channel line-up efforts. Comparison of initial pictures with succeeding results provided an accurate indication of progress.

Admittedly, while this system has not provided broadcast quality video transmission on all channels simultaneously and consistently, it compares favorably with transmission over a fairly long intercity system. On the other hand, impairments in quality to an extent acceptable in community antenna distribution systems could not have been tolerated, owing to the more rigid requirements set by the extensive use of large-screen television projectors.

Even though the over-all video signal-to-noise ratio has been satisfactory, low-frequency noise has been troublesome, particularly where it has keyed the vertical oscillator in the video-tape machines and caused vertical rolls. In most cases, this resulted from low r-f levels caused by moisture in cables. Also, power distribution system electric faults have introduced impulse noise.

Radio-frequency interference has been fortuitous and therefore difficult to locate and eliminate. The r-f field in this area varies in intensity and frequency in accordance with the number and types of radar and r-f test sets in operation. The most serious source of r-f interference is caused by intermodulation products generated within the various pieces of

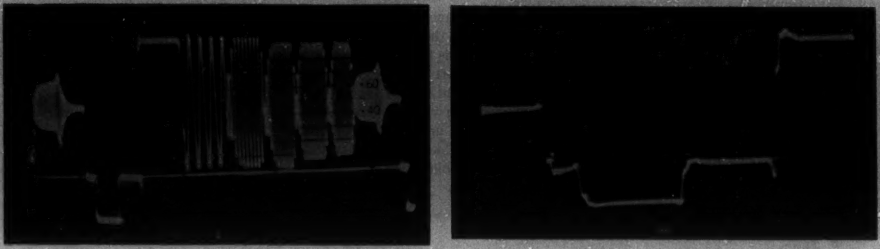


Fig. 7. Typical multiburst pattern with band-pass filter inserted, calibrated 1 volt peak to peak (A). Blanking interval of (A) expanded to show ringing (B)

equipment and falling within the pass band of a particular channel. This has proved to be a particular problem with a combination of channels 4 and sub-7, causing r-f patterns to occur in channel 6. Although it has been impossible to eliminate completely this interference, close co-ordination of levels at common amplifier points has been helpful to keep it within acceptable limits.

Ringing has been particularly troublesome throughout the entire operation, especially on projected pictures. It has been found that r-f band-pass filters and video receivers are the major contributors to ringing problems. The test layout shown in Fig. 6 was set up in the maintenance shop for the purpose of studying the causes of ringing. The response pattern shown in Fig. 7(A) is the result of passing a multiburst signal through the transmitter and receiver, with an r-f band-pass filter inserted. In this case, both the receiver and filter had been tuned for optimum results. An expanded view of the blanking interval is shown in Fig. 7(B). Here, the presence of ringing is obvious. It was also noted that the loss of setup in the multiburst pattern was because of the presence of the filter. Later it was discovered that still better results could be obtained by tuning the filters in the working circuits for which they were designed.

Cross-talk, although not as serious as noise and ringing, has been a problem. Most cases of cross-talk resulted in sync cross-over into various channels. Readable picture cross-talk, however, exists between channel sub-7 and channel 2 as a result of insufficient front-end selectivity of the channel 2 receiver, allowing the sub-7 to be fed through and accepted in the i-f (intermediate-frequency) pass band of the channel 2 receiver. While co-ordination of transmitting levels has been helpful in reducing sync cross-over it was necessary to insert a channel 2 r-f band-pass filter in front of the channel 2 receiver. Addition of this filter has not eliminated the cross-talk completely, but has reduced it to the extent that no longer is a readable picture evident. Cross-talk is now responsible for only slight loss of definition in channel 2 pictures.

The usable frequency response of any channel is limited by high frequency roll-off such that ringing will be at a minimum. In most cases, better detail can be obtained with gradual roll-off in the interest of reducing ringing. The multiburst pattern shown in Fig. 7 (A) indicates optimum roll-off for an acceptable channel.

Stability of operation has been of major concern throughout this project. Any of the operating deviations discussed have been subject to change from hour to hour and day to day. R-f interference evident in channel 2 one day might clear and then be present on some other channel the next day. Channels aligned for optimum transmission one afternoon might be unacceptable the following morning. This is caused in part by temperature differences at the various equipment locations and temperature changes from hour to hour.

Transmitters and channel-to-subchannel converters are installed in aluminum body vans in addition to the remote TV equipment. Even though these vans are equipped with exhaust fans, the temperature level becomes quite high on warm days. Line amplifiers are installed in pole-mounted boxes painted white for heat dissipation. The temperature in these boxes never reaches the temperature in the vans, in spite of the absence of forced ventilation. Subchannel-to-channel converters and video receivers are located in the master-control room which is air-conditioned to accommodate all heat generated in the area.

Master oscillators in the transmitter, although crystal controlled, are not equipped with ovens and are subject to some frequency drift with temperature changes. Master oscillators in the subchannel converters are not crystal controlled and are subject to intolerable frequency shift with temperature changes. It is obvious that frequency drift in components in the various temperature zones will not be linear and receivers will not necessarily track their respective transmitters.

Equipment maintenance, initially a serious problem because of lack of experienced technicians and comprehensive instruction bulletins, has become a routine matter. Personnel selected to operate the system have become quite proficient in maintaining and aligning all components involved.

Cost of maintenance for a system of this type is still a matter of conjecture. However, a routine has been established by which all time spent maintaining and operating this system is coded and logged to the various phases, such as, system line-up, channel line-up, transmitter line-up, transmitter repair, etc. In addition, individual maintenance records are maintained on each unit of equipment. This record shows trouble indication, trouble location, specific vacuum-tube usage, and time required on each maintenance effort. It is felt that these records will be of great value in estimating cost of maintenance of future systems and of particular types of equipment.

Conclusions

It is becoming increasingly evident that aerial-cable closed-circuit TV utilizing r-f distribution techniques has a definite place in both industrial and educational fields. In a number of applications TV service can be provided in this manner which could not be provided economically or practicably in any other way. For example, to have provided the Army with a base-band system in this case, would have required a separate underground cable from each of the transmitting points with separate amplifiers for each of these cables at a cost which would have been almost double the cost of the existing system. Reasonably good service can, however, be provided quite economically if strict maintenance procedures are utilized throughout the system. To permit expansion in this field, it is our belief that continued effort should be applied toward improvement in the design and manufacture of the various components which go to make up a system of this type, in order to achieve uniformity, stability, and dependability.

References

1. A TALK-BACK SYSTEM FOR EDUCATIONAL TELEVISION, R. C. Norton, J. B. Davidson. *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 78, May 1959, pp. 209–14.
2. A TRANSMISSION DESIGN FOR CLOSED CIRCUIT EDUCATIONAL TELEVISION, W. L. Wall. *Ibid.*, Sept., pp. 340–46.
3. CLOSED-CIRCUIT NETWORK FOR EDUCATIONAL TELEVISION, W. C. Warman. *Ibid.*, vol. 76, Sept. 1957, pp. 440–44.

An Omnidirectional Circular Antenna Array Excited Parasitically by a Central Driven Element

H. P. NEFF
NONMEMBER AIEE

J. D. TILLMAN
ASSOCIATE MEMBER AIEE

THERE IS a continuing practical need for antennas which radiate a pattern which is omnidirectional in azimuth, but with a specified shape in the vertical direction. One example is that of a broadcast station, where the radiated power should be concentrated along the ground. Conventional solutions of this problem make use of an array of colinear elements, or of sectionalized elements, or of top-loaded antennas. The height of the antenna is often of the order of a half-wavelength. Page[1] proposed that a circular array be used for this purpose, and showed that an array of several radiators distributed about the circumference of a circle, and with an element at the center, could be made to radiate an omnidirectional pattern in azimuth, with a null at any prescribed vertical angle. In the antenna described by Page all of the elements are driven, which entails complicated matching and phasing networks. It is the purpose of this paper to describe an antenna, similar in geometry to that of Page, in which only the center element is driven. The remaining elements are excited by mutual coupling from the center element, and the desired null in the vertical pattern can be controlled either by varying the length of the parasites, or by inserting a tuning reactance at the center of the parasites.

Derivation of the Expression for the Radiation Pattern

The antenna described in this paper is shown in Fig. 1, and consists of m identical, parallel parasitic elements, which are spaced symmetrically about the circumference of a circle of radius ρ. A similar element is at the center, and is driven by a voltage V_c. From the symmetry, it is evident that equal currents will be induced in each of the parasites. Let the current in the center element be designated by I_c, and the current in each parasite by I_p. Let the antenna no. C be located at the origin of a conventional spherical co-ordinate system, oriented so that the circle lies in the plane $\theta=\pi/2$ and element no. m in the plane $\phi=0$.

The radiation from the center element depends on its half-length h_c, and will be designated by F_{ec}.

The pattern of the ring array of parasites is 2

$$F_p=mI_p\left[J_0\left(\frac{2\pi\rho}{\lambda}\sin\theta\right)+\right.$$
$$\left.2\sum_{k=1}^{\infty}j^{km}\cos km\phi\, J_{km}\left(\frac{2\pi\rho}{\lambda}\sin\theta\right)\right]F_{ep} \qquad (1)$$

In equation 1 F_{ep} is the element factor of the parasites, which depends upon their half-length h_p, $J_n(x)$ is the Bessel function of the first kind of order n, and λ is the wavelength. All of the terms in the summation can be made negligible by choosing m so that

$$J_0\left(\frac{2\pi\rho}{\lambda}\right)>>J_m\left(\frac{2\pi\rho}{\lambda}\right) \qquad (2)$$

For a circumference of the order of 2.4 which will be shown to be desirable, $m=6$ makes the pattern

$$F_p=mI_pF_{ep}J_0\left(\frac{2\pi\rho}{\lambda}\sin\theta\right) \qquad (3)$$

to within less than 0.5 decibel for all directions θ or ϕ. This choice is discussed in reference 2. For the remainder of the discussion m will be assumed to be large enough to satisfy equation 2.

The pattern of the entire array is the sum of the patterns of the center element and of the ring. Then

$$F=I_cF_{ec}+mI_pF_{ep}J_0\left(\frac{2\pi\rho}{\lambda}\sin\theta\right)$$
$$=I_cF_{ec}\left[1+\frac{mI_p}{I_c}\frac{F_{ep}}{F_{ec}}J_0\left(\frac{2\pi\rho}{\lambda}\sin\theta\right)\right] \qquad (4)$$

Equation 4 gives the radiation pattern of the array. The pattern is omnidirectional in azimuth, since there is no dependence on ϕ. The pattern is dependent on θ in two ways: the element factors are functions of θ, and the Bessel function has a function of θ as its argument. The element factor of a thin dipole is

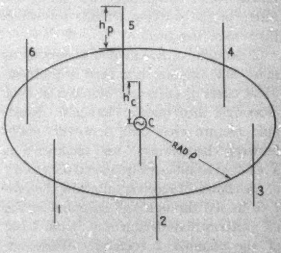


Fig. 1. Antenna array consisting of six parasites equally spaced on the circumference of a circle and a central driven element

---

Paper 60-11, recommended by the AIEE Broadcasting Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., January 31–February 5, 1960. Manuscript submitted May 21, 1959; made available for printing September 28, 1959.

H. P. NEFF and J. D. TILLMAN are with the University of Tennessee, Knoxville, Tenn.

The research reported in this paper has been supported by the Air Force Cambridge Research Center under Contract *AF 19(604)-4967*.

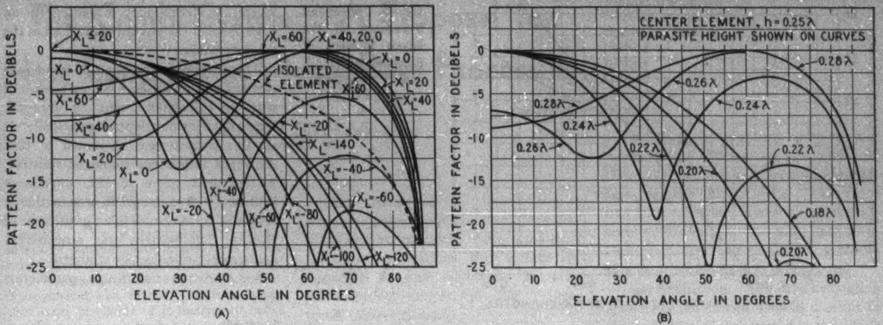


Fig. 2. Vertical patterns obtainable from the array of Fig. 1 by varying

A—Parasite termination
B—Parasite height

$$F_e=\frac{\cos\left(\frac{2\pi h}{\lambda}\cos\theta\right)-\cos\frac{2\pi h}{\lambda}}{\sin\frac{2\pi h}{\lambda}\sin\theta} \tag{5}$$

Inspection of equation 5 indicates that the ratio F_{ep}/F_{ec} is always real, and provided h_p is not greatly different from h_c, it is very nearly equal to one. It follows from equation 4 that a null exists at an angle θ_n such that

$$J_0\left(\frac{2\pi\rho}{\lambda}\sin\theta_n\right)=-\frac{I_c}{mI_p}\frac{F_{ec}}{F_{ep}} \tag{6}$$

and that for such a null angle to be real requires I_c/I_p to be real. If I_c/I_p is complex, a quasinull exists at the real angle θ_n given by

$$J_0\left(\frac{2\pi\rho}{\lambda}\sin\theta_n\right)=-\Re\left(\frac{I_c}{mI_p}\right)\frac{F_{ec}}{F_{ep}} \tag{7}$$

and the residual field in the direction of this quasinull is proportional to

$$\Im\left(\frac{I_c}{mI_p}\right)$$

Derivation of Expression for Currents and Impedance

It is evident that the performance of the array depends on the ratio I_c/I_p. This ratio depends on self- and mutual impedances of all of the elements. Let each parasite be terminated in a load impedance Z_L so that the voltage at the center of each parasite is $-I_pZ_L$. As in any linear system one can write

$$\left.\begin{aligned} V_c&=I_cZ_{cc}+I_1Z_{1c}+\ldots+I_mZ_{mc}\\ -I_1Z_L&=I_cZ_{1c}+I_1Z_{11}+\ldots+I_mZ_{1m}\\ &\vdots\\ -I_mZ_L&=I_cZ_{mc}+I_1Z_{1m}+\ldots+I_mZ_{mm}\end{aligned}\right\} \tag{8}$$

The symbols in equations 8 have the conventional meanings: that is, if $i\neq j$, Z_{ij} is the mutual impedance between elements i and j, and if $i=j$, Z_{ii} is the self-impedance of element i. Now from symmetry, $I_1=I_2=\ldots I_m=I_p$, $Z_{1c}=Z_{2c}=\ldots Z_{mc}$, and $Z_{12}=Z_{1m}$, $Z_{13}=Z_{1(m-1)}$, ... Then equations 8 reduce to two equations. These are

$$\left.\begin{aligned} V_c&=I_cZ_{cc}+mI_pZ_{1c}\\ 0&=I_cZ_{1c}+I_p(Z_{11}+Z_{12}+\ldots Z_{1m}+Z_L)\end{aligned}\right\} \tag{9}$$

The impedance of ${}_0Z=Z_{11}+Z_{12}+\ldots+Z_{1m}$ is the terminal impedance of each element in the circular array with all of the elements driven in phase.[3] It is called the zero-sequence impedance, and may be calculated from mutual impedance data.[4,5] If $h_p=h_c$, these data also give Z_{1c}. If $h_p\neq h_c$, the expression developed by Cox applies.[6] Equations 9 now become

$$\left.\begin{aligned} V_c&=I_cZ_{cc}+mI_pZ_{1c}\\ 0&=I_{c1}Z_c+I_p({}_0Z+Z_L)\end{aligned}\right\} \tag{10}$$

From the first equation, the impedance of the driven element is

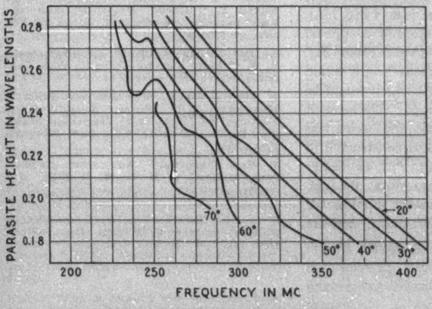


Fig. 3. Measured null angles for the array of Fig. 1

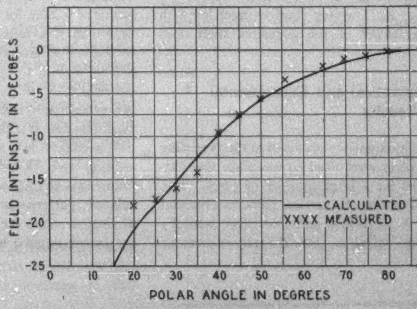


Fig. 4. Measured and calculated patterns for the array of Fig. 1 with $h_p=0.20\lambda$, and $\rho=0.383\lambda$

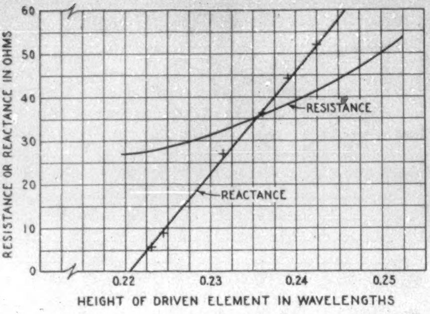


Fig. 5. Measured impedance of the driven element of the array of Fig. 1 with $h_p=0.20\lambda$, and $\rho=0.383\lambda$

$$Z_c=\frac{V_c}{I_c}=Z_{cc}+\frac{mI_p}{I_c}Z_{1c} \tag{11}$$

and from the second

$$\frac{I_c}{I_p}=-\frac{{}_0Z+Z_L}{Z_{1c}} \tag{12}$$

It is this ratio upon which the location of the null was shown to depend. This ratio depends on the radius of the ring, the element heights, and on Z_L. If there is to be no loss in Z_L, this impedance must be a pure reactance jX_L. A fortunate coincidence occurs in the numerical values of ${}_0Z$ and Z_{1c} which makes the ratio I_c/I_p nearly real if ρ is about 0.4λ. At this radius $|{}_0Z|$ goes through a minimum, and has only a small real part. Also, Z_{1c} is nearly a pure reactance at this spacing. Then, with ${}_0Z={}_0R+j{}_0X$

$$\frac{I_c}{I_p}=-\frac{{}_0R+j({}_0X+X_L)}{Z_{1c}} \tag{13}$$

Since both numerator and denominator are nearly pure imaginaries, their ratio is nearly real, and according to equation 7 a deep quasinull will exist. The angle θ_n of this quasinull can be controlled either by varying X_L, or by varying the height of the parasites, which varies both ${}_0X$ and Z_{1c}. Z_{1c}, however, changes much less with the height of the parasites than does ${}_0X$.

Theoretical and Experimental Results

The equations obtained in the preceding sections were used to calculate the patterns for a large number of cases. Fig. 2(A) shows the vertical patterns obtained for $h_c=h_p=\lambda/4$, and $\rho=0.388\lambda$, with X_L as the variable. Fig. 2(B) shows a similar set of patterns where $h_c=0.25\lambda$, $\rho=0.383\lambda$, $X_L=0$, and h_p is the variable. In both cases it is seen that quasinulls can be obtained at any desired elevation angle.

To obtain experimental confirmation, a model was constructed with the dimensions given above at 300 megacycles. Fig. 3 shows experimental null angles as a function of frequency. The irregularities in the curves for null angles of 60 and 70 degrees are caused by the difficulty in precisely locating a null in a region where the field intensity is low on either side of the null. The thickness of the elements in the model which affects the self reactance accounts for slight differences in the measured and calculated position of the null.

Typical Design

As an example, an array of the type which could be used by a broadcast station to reduce fading will be designed. For this purpose, the elements should be monopoles, driven against ground, instead of dipoles. A quasinull at a polar angle of about 30 degrees (elevation angle 60 degrees) meets this need. Using a radius of 0.383λ, for the case of Fig. 2(A), ($h_p=0.25\lambda$), a termination of $X={}_L-60$ to $X_L=-80$ would serve. From Fig. 2(B) if $X_L=0$, the parasite height would between 0.2λ and 0.21λ, which would be less expensive to construct. The element thickness also is a factor, and taking this into account for a model constructed of welding rod ($h/a=75$) with a center frequency at 300 megacycles, the calculated pattern is shown in Fig. 4. Measured points are also shown, and the agreement is good except at low angles, where the discrepancy is caused by an edge effect of the ground plane used in the measurements.

The measured impedance of the driven element was about $50+j85$, as compared with a calculated value of $50+j50$. Since Z_{1c} does not depend principally on h_c, this element was shortened so that the input impedance would be resistive. A measured curve of this impedance is shown in Fig. 5 and it is seen that a height of 0.22λ results in an impedance of $28+j0$. The pattern was remeasured, and was not significantly changed.

Summarizing, the design parameters are:

$h_p=0.20\lambda$

$h_c=0.22\lambda$

$\rho=0.383\lambda$

$m=6$

and this results in the pattern of Fig. 4, and a terminal impedance of the driven element of $28+j0$.

Conclusions

The method described in this paper can be used to design an array having a pattern which is omnidirectional in azimuth, and which has a null at any desired elevation angle. Given adequate data on the mutual impedances, the design is simple, and the resulting array is easy to excite and to match to a feed system. The use of parasites in the ring results in a much simpler excitation system than previously suggested designs for accomplishing the same result.

References

1. Radiation Resistance of Ring Aerials, **H. Page.** *Wireless Engineer*, London, England, vol. 25, Apr. 1948, pp. 102–09.

2. The Use of a Ring Array as a Skip-Range Antenna, **J. D. Tillman, W. T. Patton, C. E. Blakely, F. V. Schultz.** *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 43, Nov. 1955, pp. 1655–60.

3. The Theory of Linear Antennas (book), **R. W. P. King.** Harvard University Press, Cambridge, Mass., 1956, pp. 351–54.

4. Mutual Impedance of Identical Parallel Monopoles, **J. D. Tillman.** *Scientific Report, Navy Contract NCbsr-57032*, Engineering Experiment Station, University of Tennessee, Knoxville, Tenn., 1956.

5. High Frequency Models in Antenna Investigations, **G. H. Brown, R. King.** *Proceedings*, Institute of Radio Engineers, vol. 22, Apr. 1934, pp. 457–70.

6. Mutual Impedance between Antennas of Unequal Height, **C. R. Cox.** *Ibid.*, vol. 35, Nov. 1947, pp. 1367–71.

The Combi-System—A Proposal for New Concepts in Digital Data Processing

HELMUT SCHWAB
NONMEMBER AIEE

Synopsis: A new systematic organization of the field of digital data processing is proposed. This is motivated by recent extensive developments in this area. The subgroups—data acquisition, data handling, data evaluation, and data display—are distinguished, and defined; and it is shown that most problems of industrial or commercial data processing belong to the group of data handling. Data-handling problems in the past have been mostly solved by general purpose computers or systems based on computer technology. A new approach, based on basic functions of data handling is proposed. In the field of computing, basic functions are: add, subtract, multiply, and divide. In the field of circuit logic, basic functions are: and, or, not, memory, and time. The proposed basic functions for data handling are: position-selection, value-comparison, switching, memory, control, and translation. Similar to proceedings in computer technology or circuit logic, symbols are proposed for the basic data handling functions. These symbols allow a block diagram approach as well as eventually a new type of algebra in solving data handling problems. The use of these symbols is demonstrated in some practical examples. The practical building blocks for a data handling system based on the forementioned functions are discussed. To each of the functions corresponds one or a group of building blocks as selector, comparator, buffer, switch, and control unit. The advantages of this system are briefly discussed. It seems possible to combine the advantages of flexibility and ease of modification in general purpose machines with the advantages of simplicity and efficiency in special purpose equipment through a building block design of the proposed system.

THE FIELD of electronic digital data processing was first entered by scientific computers. The goal was to handle rather complex mathematical operations automatically. Their speed and volume of operation was low in comparison to today's machines. The general purpose machine was apparently the right approach, where the size was determined by the volume of operation. Soon it became apparent that commercial applications were economically more promising. In these applications the volume of operation is more important than the mathematical complexity. Fig. 1 illustrates this relationship.

This development has led to data-handling systems. Searching and filing, for example, are the main operations in the handling of patents in a patent office. These operations require the handling of a very large volume of data yet do not involve any mathematics.

An attempt has been made to handle the total area from scientific computation to large-volume data handling with increasingly large general purpose computers. It seems, however, more economical to look now for new approaches to these new areas. An approach from the data source of the data flow, as in contrast to the approach from the computer side, seems promising. The following presents an analysis of the data flow in digital processing systems.

Digital Data Processing

ORGANIZATION OF DIGITAL DATA PROCESSING

The field of digital data processing shall be subdivided into four areas:

Data Acquisition
Data Handling
Data Evaluation
Data Display (see Fig. 2)

Data Acquisition. Original information might be available in form of physical parameters, written records, or any other form. The problem of data acquisition is to transform this information into electric digital signals in the code or format required by a following data-processing system such as:

Analog-to-digital converter systems
Telemetering ground stations
Manual input devices

Data Handling. The following list of data-handling problems might best delineate the intended coverage of this area:

Data selection (reduction)
Format translation
Medium translation
Searching–reading
Sorting
Mixing
Correcting
Monitoring
Memory

None of these problems involve any mathematical operations. In addition, the output data contain the same information as the input data. The data-handling operations concern the arrangement of the data in the total data flow, their sequence, their selection or translation on other media or in other codes.

Data Evaluation. The data-evaluation area furnishes output data as a result of the evaluation of the input data and which therefore, are different from the latter. Computers and arithmetic units are typical data-evaluation systems or units.

Data Display. This area covers the equipment for conversation of the electronic data into a form which can be understood by human beings: plotters, printers, and other displays. It could also cover outputs from the data processing for any kind of action such as actuation of valves in chemical plants or control of machine tools.

Data Handling

DATA-HANDLING PROBLEMS

The organization of the field of digital data processing shall help to initiate a measure of order in this vast field and to choose the appropriate approach for each area. The approach which is good for data handling might not be ideal for data acquisition. Similarly, the computer is probably not the best approach for data handling either. A more systematic analysis of the area of data handling might be useful in arriving at a new approach.

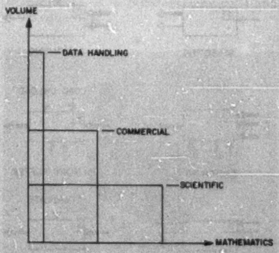


Fig. 1. Digital computers

Data-handling problems can be found in banking systems, air-traffic control, large-scale production control, instrumentation data editing, format transla-

---

Paper 60-158, recommended by the AIEE Recording and Controlling Instrumentation Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., January 31–February 5, 1960. Manuscript submitted November 23, 1959; made available for printing December 9, 1959.

HELMUT SCHWAB, formerly with the Consolidated Electrodynamics Corporation, Monrovia, Calif., is now with Applied Development Corporation, Hawthorne, Calif.

DATA ACQUISITION		DATA HANDLING		DATA EVALUATION		DATA DISPLAY
MICRO SADIC	STORAGE	EDITOR	STORAGE	COMPUTER	STORAGE	PLOTTER
TELEMETERING SYSTEM		CONTROL SYSTEM		ARITHMETIC UNITS		CONTROL
MANUAL INPUTS		X-LATOR				OUTPUTS
						PRINTER

Fig. 2. Digital data processing

tion, curve smoothening, scale correcting, process control, patent office filing, warehouse inventory, and many others.

It was already mentioned that all these applications can be reduced to a rather small number of fundamental problems. Such are:

Data selection (reduction): Systematic selection or editing of data for time selection, channel selection, and digit selection.

Format translation: Rearrangement of the electric or magnetic pulse pattern of one computer language into the "format" of another computer language. The information content is not changed in this process.

Medium translation: The transcribing of information from one storage medium to another, for instance from magnetic tape to punched cards.

Searching reading: Searching of a specific data block in a large file of information, where the block is characterized by some search digits, but not by its location. Also searching of single digits in a data block.

Sorting: For instance, alphabetic or numerical sorting of random input information. Quite important in all warehouse-type operations.

Mixing: Collating of two or more data flows each of which are sorted in the same manner.

Correcting: Mainly correcting of single digits in a data block. This operation consists of two parts, "cancel information" and "add information." The correction of a whole data block can rather be considered as a sorting and mixing operation.

Monitoring: In-limit or out-of-limit indication, very common in process and quality control, but also used in editing and other applications.

Memory: A passive operation, in contrast to the active operations already mentioned, but nevertheless very essential. It is independent from the previous operations and belongs to the area of data handling.

DATA-HANDLING FUNCTIONS

The previous paragraphs have already explained two steps in the systematic organization of data processing and data handling: the organization of the data-processing field in subareas, and the organization of data handling in basic problems. There is still one more step possible which has essential consequences for the "hardware," the definition of fundamental data-handling functions.

The basic data-handling problems can be reduced to six fundamental functions:

Selecting
Comparing
Switching
Memory
Control
Translation

These functions are basically known. It is with their systematic investigation in the area of data handling that this paper is concerned.

Position Selection. Selection or indication of a specific area in a data flow (for instance, selection of the N th block in a flow of data blocks, or of the N th digit within any data block).

Value Comparison. Comparing of two or more selected characters mainly for their numerical relation (A greater than B, A smaller than B, or A equal to B).

Data Flow Switching. Mixing of a data flow from one or more inputs or distribution to one or more outputs, or a combination of both through flow switching (gating).

Data Memory. This includes buffering as well as long time memory. A shift register might be a specific memory.

Systems Control. Control functions are needed for the co-ordination of the previous functions to complete systems. Command gating and memory are essential parts of the control, as are also clock and reset functions.

Character Translation. In certain systems it is useful to introduce the translation function. This is not required in cases of table translation (memory function) or simple rearrangement (control function). Code changes, however, are typical translations. Some scale-correction systems belong also to this group.

APPLICATIONS

The foregoing theory introduces the idea of fundamental functions for data-handling problems. It shall be mentioned that such a theory for a logical field is not new. In mathematics, it has been known to solve problems with the use of the four functions:

Add
Subtract
Multiply
Divide

In circuit logic, a similar approach yielded the five functions: And, Or, Not, Memory, and Time.

The definition of fundamental functions

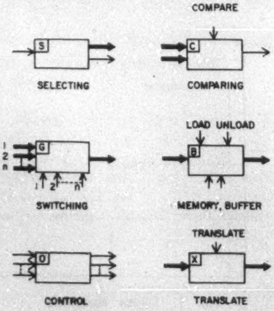


Fig. 3 (left). Data-handling symbols

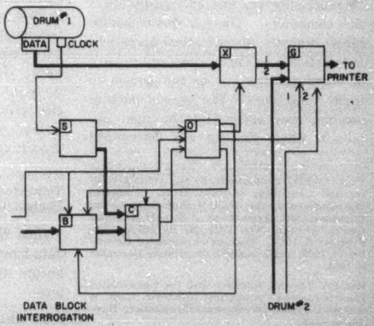


Fig. 4 (right). Data-handling application

Fig. 5 (left). Submodules

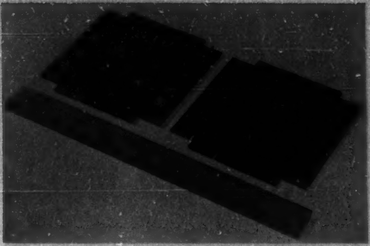


Fig. 6 (right). Printed cards

allows the introduction of symbols for their representation. Actually, there are two types of symbols, those used for block diagrams (or flow diagrams) and those used for algebraic equations. This is especially apparent in circuit logic. There, one finds the gate symbols as well as the symbols for the Boolean Algebra. One could even say that it is only the creation of symbols which allows the solution of complex problems by the method of block diagrams or algebra. The creation of data-handling symbols would, then, be an essential advantage in using the foregoing approach to data-handling problems. Fig. 3 shows block-diagram symbols for the proposed system.

Fig. 4 shows an example of a data-handling problem solved with the "Combi-System" block diagram approach. It presents in a schematic fashion a detail of a larger system. This detail is the read-out of data blocks from a magnetic drum system to an electromechanical printer as interrogated (addressed) by an external, nonsynchronous, data source.

The interrogating block address is held by the memory-function *B*. The selector-function *S* is controlled by the drum clock and, thereby is used as drum address selector. Its output is compared with the interrogating address by the comparator-function *C*. In case of a coincidence the translator-function *X* is actuated. This allows read-out of, for example, a *BCD* code if the drum had a straight binary code. The switch-function *G* permits connection of several drum systems to the same printer. The control-function *O* co-ordinates the previous functions.

It can be seen that there are two different flows of information, the data flow (heavy lines), and the command flow (light lines). This is also apparent in the symbols in Fig. 3.

The data flow is formed by four tracks plus parity check (and sign) in the *BCD* code and by six or more tracks plus parity check in the straight binary code. The transmission of true signals as well as false signals might be possible.

The command flow consists normally of one channel only. Possibly, however, a philosophy corresponding to the "standard level exchange" in Consolidated Electrodynamics Corporation's Micro-SADIC may be followed. There, each command consists of three parts:

1. The preparation signal, one clock period long prior to the command.
2. The action level for the length of the action.
3. The end level, one clock period long prior to the end.

Examples of the function of the command are:

Load, unload, recycling, or reset command for the memory function.
Comparing and decision-making command for the comparator function.
Step command for the selector function.
Gate opening for the switch function.

With these rules, Fig. 4 can be understood in the following way.

The interrogating data source delivers an address through its data flow and simultaneously a load command to the memory function. This command also instructs the control function that a new address has been delivered.

The selector unit, controlled by the clock of the drum, delivers that address on the drum which is just passing the read heads. Zero address is selected as a command to the control unit for reset purposes. The comparator unit receives the interrogation address as well as the drum address. It also receives a command telling when to effect the comparison. This command is identical with the recycle command for the memory-function. The result of the comparison is delivered to the control unit. Through the latter (or directly) it activates the translator-function and the switch function. Thereby a read-out to the printer is effected. Certain commands from the selector might be used in the control function to determine the block length.

It has already been mentioned that another set of symbols might be created for an algebraic approach to data-handling problems (similar to the Boolean algebra). This shall be explained for the problem of Fig. 4. The symbols shall be

⁚ select. $D ⁚ A$ means select data flow D in correspondence to information A.
∞ compare. $A \infty B$ means compare A and B for equality.

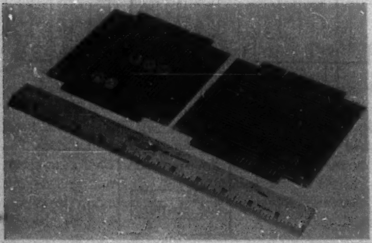


Fig. 7 (left). Logic cards

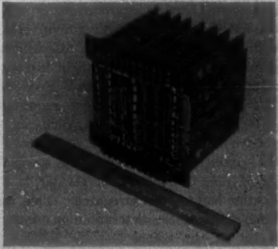


Fig. 8 (right). Submodule (complete)

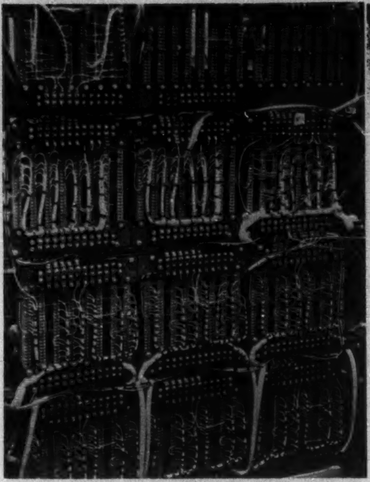


Fig. 9 (left). Combi-system model

Fig. 10 (above). Model of an integrated digital data-processing system

$*$ memory. A^* means keep A in memory.
— translate. $A-D$ means translate D in accordance with information A. This corresponds remotely to the minus sign in algebra.

If we define A as the incoming address, D as the data flow from the drum, and G as the output data flow, then the example in Fig. 4 is represented by the equation:

$$G=A^*\infty(D : A)-D$$

Another possible solution of the same problem is apparently

$$G=A : (A^*\infty D)-D$$

At this moment this equation is only a shorthand notation of the problem. It shows, however, in what direction studies have to be directed in order to establish a data-handling algebra. Its use would correspond to that of the Boolean Algebra in circuit logic.

The Combi-System has been used, for example, for the design of an air traffic control system. The problem was to give conflict monitoring, fix area monitoring, and flight history for 200 airplanes in a 400-mile-square 100,000-foot-high area with 200 fixes and 50 sectors. Updating period is 1/16 of a second, flight history evaluation period is less than 6 seconds. The conflict evaluation alone required 40,000 comparisons per scan, each comparison for three dimensions. Previously, the conflict tolerance for the specific sector (one out of 50) for the airplane location had to be interrogated. This, in turn, required the determination of the sector of the actual airplane location. Each conflict resulted in a time determination, count, and successive printout of all pertinent data.

Hardware

Building Blocks

A system has been designed which applies the foregoing theory to a series of building blocks. This system has acquired the name "Combi-System." The basic philosophy of this system is to provide a family of standard building blocks for each fundamental data-handling function. The blocks are also called "submodules." The idea is, then, to combine the standard building blocks into tailor-made systems.

The creation of the submodules is a new step in the modularization of systems. Building blocks in form of standard circuits on printed cards are widely in use. Another level of modularization is the use of standard operational units, for instance the major parts of a computer (memory unit, arithmetic unit, input-output devices, control). This type of standardization finds its limitation in the great variety of applications. The new step, as a result of the theoretical investigation of the fundamental data-handling functions, modularizes on a level which corresponds to the largest common building

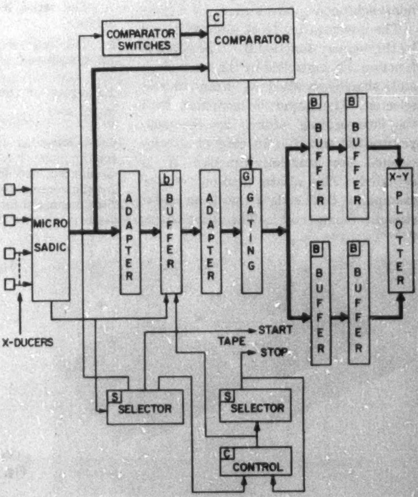


Fig. 11. Data-processing experimental system

block for all applications. These new "submodules" are formed by a certain number of standard circuits or "cards." Several submodules together can form an operational unit. Since the submodules correspond to the fundamental functions, they bring advantages for designing of systems, trouble-shooting in systems, and modification of systems. Beside that, they have the normal advantages of standard building blocks for production and maintenance.

Each submodule has provision for combining eight printed cards. The module itself consists only of the eight connectors, their wiring, and an iron frame, as shown in Fig. 5. Fig. 6 shows printed cards (one flip-flop card and one logic card), and Fig. 7 shows a general purpose logic card (crossed wire-grid type). Fig. 8 shows a submodule with printed cards inserted.

Fig. 9 shows a laboratory arrangement of submodules in a larger frame system. The interconnecting cabling in this laboratory model is achieved by card-plugs for the data flow and barrier-strip connections for the power supply and command flow. A more advanced design is being tested.

All circuits and standard cards used in the Combi-System are the same as those considered standard for Consolidated Electrodynamics MicroSADIC. Therefore, power supply and signal specifications are the same for the Combi-System as for MicroSADIC.

SYSTEM MODEL

Fig. 10 shows a feasibility model for an integrated system including data acquisition, data handling, and data display. Fig. 11 shows its block diagram.

The system model has a model rocket as test object. Transducers (potentiometers) indicate its pitch, yaw, and roll. A first oscilloscope is used as a quick-look device, displaying the pitch and yaw as an *X-Y* plot. These values are, then, sent to a MicroSADIC in unblocked data flow operation at low sampling rate, similar to a PDM link. From this data acquisition section, the data flow to a data-handling system using laboratory Combi building blocks. This section first does time editing on preset start and stop time based on time digits in the data flow. This section also translates the data flow into an International Business Machines Corporation's *704* format which might be recorded by a tape unit. The tape unit is directly controlled by the translator. The data-handling unit, finally provides a double-channel selection, again for the pitch and the yaw samples in the data blocks. The last part of the system provides data display. The electronic section also uses Combi blocks (shift register and digital-to-analog converter). The display is again in the form of an *X-Y* plot. At present, a cathode-ray tube is used for the final display. Later, a photographic process or the like might be more appropriate.

This feasibility model shall prove how data handling fits into the total data-processing system and how essentially the same hardware can be used in all parts of an integrated system, from data acquisition to data display.

Conclusions

The Combi-System has different aspects. One is the more philosophical aspect, treating the data-handling theory. The result is the systematic definition of the fundamental functions, select, compare, switch, control, memory, and translate. It also includes the introduction of symbols for these functions, allowing the solution of data-handling problems in simple block diagrams, and possibly with specific algebraic methods.

Another aspect is the hardware, created as a consequence of the theory. This includes the fundamental Combi-System building-block families: comparator, selector, memory, control, and translator. New ideas for framing, cabling, and front panels are connected to these new building blocks. There are no essentially new technologies used in this approach. The hardware utilizes standard circuitry and cards from MicroSADIC.

The previously defined data-handling functions and circuits are basically known. The contribution of this paper shall be their systematic investigation and definition under the aspect of general use in data handling. This simplifies considerably the design and manufacturing of data-handling systems. The Combi-System is a new systematic approach to the field of data handling.

In this new approach, the foregoing theory allows a practical way of solving even complex data-handling problems (for instance, the air traffic control problem) and the new hardware arrangement allows an economical design of such systems with already available material.

This report has presented new concepts in organizing the field of digital data processing. Especially for data handling, the attempt has been to show new ways for problem solution using Combi-System theory and standard hardware in fundamental functions for using standard building blocks.

A Precision, Continuous Voltage Reference for Industrial Recorders

PAUL B. ROBINSON
MEMBER AIEE

A NEW PRECISION voltage reference for use in industrial null-balancing recorders is described. The reference employs a Zener diode as the voltage element and replaces the conventional dry cell, standard cell, and standardizing mechanism used in most recorders. The reference is currently being supplied as standard equipment in the new General Electric strip chart recorder.

The Need for a Voltage Reference

In the past several years, a number of attempts have been made to eliminate the standard cell and dry battery that have been standard equipment in most industrial null-balancing potentiometric recorders. While these voltaic cells have rendered long and dependable service, they have also represented a service expense and inconvenience to the user, who has had to replace the dry cell periodically, hopefully, but not necessarily, before the useful life of the dry cell has been exceeded. The standard cells have also limited the temperature to which a recorder could be exposed, since the cells freeze only a few degrees below the freezing point of water. In cases where excessively low temperatures were unavoidable, heaters had to be added to the recorder or the standard cell located re-

---

Paper 60-114, recommended by the AIEE Recording and Controlling Instrumentation Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., January 31–February 5, 1960. Manuscript submitted November 23, 1959; made available for printing November 27, 1959.

PAUL B. ROBINSON is with the General Electric Company, Lynn, Mass.

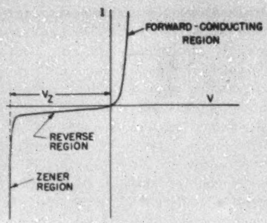


Fig. 1. Volt-ampere characteristics of Zener diodes

mote from the recorder to prevent freezing.

The necessity of periodically standardizing the dry cell against the standard cell has meant momentary interruption of the measurement of the process variable and the loss of control or alarm functions. This loss is, of course, for brief intervals only. But regardless of how brief, the standardizing mechanism with its clutches, cams, and switches, has been an unavoidable part of every recorder using the cells.

The standardizing rheostat also represents a problem with instruments having highly suppressed scales, since each convolution of the rheostat winding begins to represent a noticeable shift in the instrument's calibration. Thus, the standardizing adjustment no longer appears continuous but is a series of discrete steps. The result is apt to be shifts in instrument calibration following each standardizing period.

For these reasons, a replacement for the standard cell-dry cell combination has long been recognized as desirable. Needed is a voltage reference that is at once simple, virtually maintenance-free and above all, sufficiently precise to be used as a reference in the 1/4% class of industrial null-balancing recorders where voltaic cells are now used.

This paper describes a reference which, it is believed, meets these requirements. The reference is currently being applied in the new General Electric types *HF* and *HG* recorders, and has been used in special applications with the type *HE* round chart recorder.

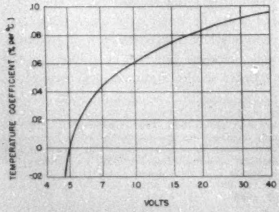


Fig. 2. Zener breakdown voltage

Zener Diodes

Volt-Ampere Characteristics

The principal element in the reference to be described is a Zener diode. These diodes have been described in detail in published literature and are simply silicon junction diodes. They exhibit the volt-ampere characteristic shown in Fig. 1, which is very similar to the volt-ampere characteristic for other semiconducting diodes except for the breakdown or Zener region. In this region, the voltage across the diode is nearly independent of the current flowing through the diode, a trait that immediately suggests the possibility of using the diode as a reference.

Comparison of Zener Diodes with Gas Regulator Tubes

The similarity between the volt-ampere curve of the silicon diode in the Zener region and the volt-ampere curve of a gas regulator tube is obvious. The diode has two immediate advantages over the regulator tube, however. One is that the diode does not need a voltage in excess of its Zener voltage to conduct in the breakdown region as does a gas tube. Thus, "starting" the diode is no problem and need not be a part of the calculation of circuit parameters as it must with the gas regulators. The second advantage is that there are no discontinuities in the volt-ampere curve in the diode's Zener region as there are in gas regulators. These discontinuities are caused by sudden changes in the ionization pattern within the gas tube as current varies. The diodes seem to have no comparable phenomenon, for the slight effect of diode current on Zener voltage is a continuous and dependable one.

Temperature Characteristics of Zener Diodes

Silicon diodes are available with a wide range of Zener voltages, but the selection is limited when the need is for a precision reference. This is because the breakdown voltage is a definite function of temperature, with the magnitude of the temperature effect being a function of the breakdown voltage. This relationship is shown in Fig. 2, which indicates that for reference work, where a low temperature coefficient of voltage is necessary, relatively low voltage units are called for.

Experience has shown that while the lowest temperature coefficient appears to be around 5 volts, the predictability of this coefficient from diode to diode is not so good at 5 volts as it is at 7 volts. Thus, it has proved better to work with the 7-volt diodes and compensate them for their positive but reasonably dependable temperature coefficient.

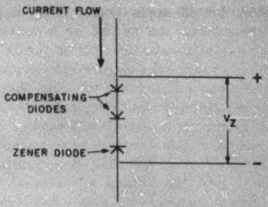


Fig. 3. Circuit of temperature-compensated Zener diodes

Temperature Compensation

Compensation in commercially available reference diodes is accomplished by connecting two additional silicon diodes in series with the Zener diode, but in the forward direction, as shown in Fig. 3. The forward-connected diodes have negative temperature coefficients of voltage, their voltage drop decreasing with increasing temperature. The magnitude of this coefficient is about half that of the Zener diode's positive temperature coefficient, and, by proper selection and matching, the two forward diodes can be made to compensate the Zener diode exactly. The result is that units with temperature coefficients very close to zero are obtainable, and this over a wide temperature range.

The voltage drop in the compensating diodes accounts for the unit's voltage rating being above 7 volts. The *1N430* is rated at 8.0 to 8.8 volts.

Dynamic Resistance

While the diode voltage in the breakdown region is very nearly independent of current, it is not completely so, there being a slight slope to the volt-ampere characteristic. If the diode's Zener voltage is considered equivalent to a battery,

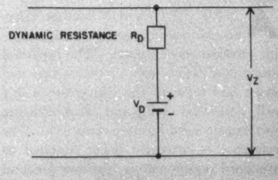


Fig. 4. Equivalent circuit of Zener diode

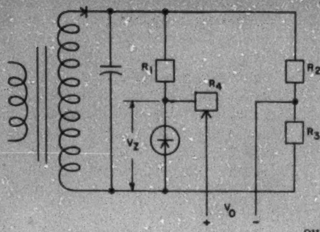
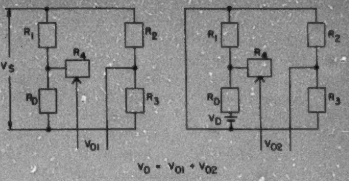


Fig. 5 (left). Circuit of precision reference

Fig. 6 (right). Equivalent circuits of reference resulting from superposition



the slope of the characteristic curve can be represented by an equivalent resistance in series with the battery, as shown in Fig. 4. Another inherent advantage of the Zener diode over the gas regulator tubes is that this equivalent, or dynamic resistance, is an order of magnitude lower in the diode, being about 15 ohms in a temperature-compensated reference diode.

The Precision Reference Unit

DESIGN CRITERIA

The design of the General Electric precision voltage reference started with a commercially available temperature-compensated Zener diode, the *1N430*. This diode is rated at 8 to 8.8 volts and has a dynamic resistance of 15 ohms at 10 ma (milliamperes). Its temperature coefficient is no greater than ±0.002%/C (degrees centigrade).

Given a temperature-compensated diode then, the principal problem is to eliminate the effect of the diode's dynamic resistance. Two ways of doing this are available. One is to provide a buffer for the diode so that it sees virtually no line voltage variations. This can be done by using a constant-voltage transformer or a higher voltage and uncompensated Zener diode in the reference diode's power supply. In this way, line voltage variations of ±10% can be reduced to ±1% or less, a wholly reasonable value. If diode current is 10 ma, a 1% variation would be 0.1 ma. This change across the 15-ohm dynamic resistance is 1.5 millivolts out of 8 volts, or 0.02%.

Another way of eliminating the effect of dynamic resistance, however, and the one used in the reference described, is to compensate for the resistance with a bridge network. The circuit is shown in Fig. 5.

Resistor R_1 and the Zener diode form one leg of the compensating bridge, and R_2 and R_3 the other leg. Replacing the Zener diode with its equivalent circuit (Fig. 4) and then analyzing the circuit by superposition gives the two equivalent circuits shown in Fig. 6. Here the net output voltage, V_o, is the sum of the two voltages, V_{o1} and V_{o2}. Obviously, V_{o1} is the line voltage-sensitive component, while V_{o2} is independent of line voltage. V_{o1} can be brought to zero by balancing the Wheatstone bridge, thus eliminating line voltage effect. The bridge is balanced by letting $R_1/R_D=R_2/R_3$.

Resistor R_1 serves the additional purpose of limiting the current through the diode to the desired 10 ma. The supply voltage, V_s, and thus R_1, want to have as high a value as practical so that $V_s-V_z/V_s \rightarrow 1$, where V_z is the Zener diode voltage. In this way, a given percentage change in V_s, caused by a line voltage change, will produce only a slightly greater percentage change in diode current. If, however, as an example, $V_s-V_z/V_s=0.5$, the percentage change in the diode would be doubled, making the compensation for dynamic resistance more difficult.

The success of the bridge-balancing compensation described above is only as good as the assumption that R_D and V_D are independent of diode current. Neither is entirely constant, of course, but the assumption is good enough so that ±10% line voltage changes produce within ±0.03% output voltage changes.

One of the problems involved in designing the bridge resistors is measuring the dynamic resistance, R_D. The fastest method is to impose the 10-ma direct current on the diode, then to modulate this direct current with a 60-cycle current of known magnitude. Measuring the alternating voltage across the diode and dividing by the alternating current gives a value for R_D. However, resistors R_2 and R_3, computed from this value, do not give the best compensation and it was discovered that the 60-cycle method does not include a slow (less than a second, however) thermal effect. Therefore, a more dependable way of measuring R_D is to impress two values of current in the region of interest and measure the two corresponding voltages. The resulting ΔE divided by the ΔI gives a good measure of R_D.

The reference unit is designed to supply a fixed current of 2 ma to a known load at 7.5 volts. It will not regulate with the same precision with variable output currents. To adjust for the variation in diode voltage from unit to unit, resistor R_4 is added. This is adjusted until the output voltage is the desired 7.5 volts at the rated 2 ma.

Life Test Data

The *1N430* diode has been extensively tested to determine its suitability as a reference on the basis of age and various environmental influences. The following tests were made:

1. Diode voltage as a function of time
2. Dynamic resistance as a function of time
3. Temperature coefficient as a function of time
4. Extreme temperature exposure
5. Thermal shock
6. Repeatability
7. Humidity
8. Vibration
9. High potential

1. The life tests were conducted on a large sample of diodes, which were operated at approximately 10 ma continuously. The diode voltages were checked at precisely 10 ma periodically. Most diodes exhibited a tendency to increase their output voltage slowly with time. When deviation was plotted against the logarithm of time, the resulting curve was roughly a straight line. With failure arbitrarily defined as 0.05% change of breakdown voltage, the best least-square line through the data predicted an average life of 71 years. The scatter of the data indicated 95% confidence that no more than 5% of the diodes would fail before 2.7 years.

One of the most difficult parts of the life tests, and all other tests, for that matter, was in setting up test equipment with stability good enough to give the needed accuracy. Since the tests were run beyond 2 years, the requirements were rather severe.

2. It was important to test the stability of the diodes' dynamic resistance because of the bridge circuit used to bal-

ance its effect on the output of the reference unit. The variation of the dynamic resistance with time was found to be small enough to be completely neglected.

3. Likewise, the temperature coefficient showed no tendency to vary with time.

4. The most severe test on the diodes was the extreme temperature exposure. The diodes were exposed to −55 C and 100 C, with checks at 25 C between exposures. This indicated a slight hysteresis effect on diode output, with the low temperature exposure increasing voltage at 25 C and the high temperature decreasing voltage. The cause of this effect is unknown, though it may be due to adsorption and desorption of impurities on the silicon junctions from the surrounding material.

This severe test produced one of the few failures, as one diode shifted 0.18% from its original breakdown voltage.

5. Five of the *1N430* diodes were subjected to 14,800 thermal shocks over a 6-day period. The thermal shock consisted of pulsing the diodes at their rated power dissipation, which is well above their normal operating level. The effect here was negligible, indicating that the diodes are conservatively rated.

6. In order to test repeatability, diodes were switched on and off over 600 cycles, the voltage during each "on" period being recorded automatically. Results indicated that the diode's ability to produce the same voltage repeatedly is comparable to that of a standard cell.

7. Five diodes were tested in high humidity as outlined in military specification *MIL E-5272A*. The results showed a maximum voltage variation 0.012% and average of 0.010%. Humidity evidently has negligible effect.

8. Diodes were subjected to vibration in three planes with the following frequencies and magnitudes:

Frequency cps	Total Travel or Acceleration
5 to 10	40 mils
10 to 75	16 mils
75 to 500	5 g*

* Gravitational acceleration.

These tests indicated no effect on diode voltage.

However, it was discovered during the vibration tests that the diode voltages are slightly sensitive to diode position. A change of 0.018% resulted from a 90-degree change in orientation. The presumed reason for this was the change in heating effect caused by the flow of the silicone oil in which these diode junctions were immersed. As the diodes are applied in a recorder, the position influence is of no consequence, as the diode's orientation is fixed.

9. A 500-volt potential was applied to five diodes from case to one terminal. No failures were found, and diode voltage was virtually unaffected.

Conclusions

The results of these tests indicated that the diodes meet the manufacturer's specifications and that they further appear to be stable over long periods. They also withstand many severe environmental conditions. On the basis of these data, the diode has been used in the precision reference described in this paper, which in turn is being supplied as a part of the General Electric Company's strip chart recorder.

Vol
Pt.
196

Thermal Voltage Converters for Accurate Voltage Measurements to 30 Megacycles Per Second

F. L. HERMACH
MEMBER AIEE

E. S. WILLIAMS
NONMEMBER AIEE

Synopsis: Thermal voltage converters, each consisting of a resistor in series with a thermoelement in a coaxial line, have been developed for measurements of rms voltages of 1 to 200 volts at frequencies from 3 cps (cycles per second) to 30 mc (megacycles). An accuracy of 0.1% or better may be obtained by a-c–d-c transfer techniques up to at least 10 mc and 0.2% at 30 mc.

A THERMAL VOLTAGE converter, according to ASA Standards,[1] is a thermoelement of low current input rating with an associated series impedance or transformer, such that the emf (electromotive force) developed at the output terminals gives a measure of the voltage applied to the input terminals. Thermal voltage converters containing thermoelements in series with wire-wound resistors can be used to make highly accurate voltage measurements at audio and ultrasonic frequencies.[2,3] The frequency range of such voltage converters is limited primarily by the residual reactances of their wire-wound resistors, but tests have indicated that for some of them good performance might be expected to frequencies approaching 1 mc.

This paper describes thermal voltage converters of low and computable reactance that have been developed at the National Bureau of Standards to meet the need for determining the frequency limit of such instruments. Single-range converters with deposited-carbon resistors have been constructed with ranges of 1 to 200 volts. Each has a frequency influence less than 0.1% to 10 mc and less than 0.4% at 30 mc. These rms voltage converters may also be used to calibrate 1/2% r-f (radio-frequency) thermocouple voltmeters which are now commercially available, and, with sine wave generators, could be used to calibrate electronic voltmeters as well.

Description and Construction of Converters

The r-f thermal voltage converters make use of the transfer principle,[2] in which a direct voltage is substituted for the alternating voltage to be measured. The direct voltage is adjusted to give the same output emf of the thermoelement that was obtained with the alternating

---

Paper 60-135, recommended by the AIEE Indicating and Integrating Instruments Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., January 31–February 5, 1960. Manuscript submitted October 30, 1959; made available for printing December 9, 1959.

F. L. Hermach and E. S. Williams are with the National Bureau of Standards, Washington, D. C.

The authors wish to acknowledge the help of R. Richardson and J. Scofield, who checked the equations and carried out the computations which made possible Fig. 4 and the last column of Table II, and of J. E. Hill, who assisted with some of the measurements. Thanks are also extended to M. C. Selby and L. Behrent for their careful calibration of the 1-volt converter.

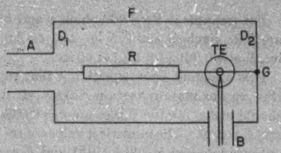


Fig. 1. Thermal voltage converter

R=Film-type resistor 1/8 in. diameter, 2 in. long
TE=Thermoelement
A=Coaxial input connector, GR874PB
B=Two-pin output connector, AN10SL
D_1, D_2=Circular brass disks, 3/16 in. thick, secured to F with 4 2-56 brass screws
F=Cylindrical brass shield, 2 in. diameter, 4.25 in. long
G=Cross wire at center of D_2 (see text)
Note: Plane of thermocouple leads at right angles to axis of F

voltage applied, and is measured with a potentiometer or other suitable means. Thus only good short-time stability, high precision of reading, and small known frequency influence are required of the transfer standard.

Each converter consists of a cylindrical deposited-carbon resistor in series with a uhf (ultrahigh frequency) thermoelement having a short straight heater in line with its supports, coaxially mounted in a brass cylinder, as shown schematically in Fig. 1. The residual reactances are much smaller than those of wire-wound resistors, and can be computed at least approximately, so that the frequency errors can be estimated. The frequency errors of the uhf thermoelements as current converters are believed to be small to over 100 mc.

The dimensions of the converters and additional data are given in the caption of Fig. 1. Each converter is assembled by first soldering the resistor, *R*, to the thermoelement, *TE*, and to the coaxial input connector, *A*, which is mounted in disk, D_1. This disc is then fastened to the cylinder, *F*, and the two thermocouple leads are soldered to the 2-pin output connector, *B*. Disk D_2, which has a small hole with a cross wire at *G*, is then fastened to *F* and the free end of the heater of the thermoelement is soldered to the cross wire. For clarity both thermocouple leads are shown in the figure. However the plane of the thermocouple leads and the pins of the output connector are actually at right angles to the axis of the cylinder so that minimum emf is induced in the output circuit. In use *F* must be grounded through the input connector.

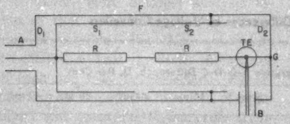


Fig. 2. Voltage converter with compensating shields

S_1=Cylindrical brass shield, 2 in. diameter with one brass end-plate
S_2=Cylindrical brass shield, 2 in. diameter both ends open, axially adjustable (see text)
Other symbols same as Fig.1 except that F is 2.5 in. diameter and 7.5 in. long
Note: Plane of thermocouple leads at right angles to axis of F

The 100- and 200-volt converters contain two resistors and two cylindrical inner shields, S_1 and S_2, as shown in Fig. 2, to reduce the effect of the distributed capacitance from the resistors to the outer cylinder. This considerably extends their frequency range. Shield S_2 is centered in *F* by a narrow ring, and is fastened to *F* by three equally spaced screws. Slots in *F* make axial adjustment of S_2 possible. The end-plate of S_1 is soldered to the input lead and to the resistor in assembly and this shield is fastened to D_1 by small insulators. The assembly is otherwise similar to that described for the other converters. The 50-volt converter has a single resistor and a fixed inner shield like S_1 in a cylinder of the dimensions given in the caption of Fig. 1.

Additional information on the voltage converters is given in the first three columns of Table I. The commercially available deposited carbon resistors are 1/8 inch in diameter and 2 inches long, and are vacuum sealed within a 3/16-inch-diameter glass cylinder. Some resistors of a different construction showed unaccountable a-c–d-c errors above 10 mc when used in earlier converters. The 5- and 10-ma uhf thermoelements, each with colinear supports and heater wire, are also commercially available, with an output voltage of about 7 mv (millivolts) at rated current. Each is in an acorn shaped evacuated glass bulb, and has a small bead between the heater and hot junction of the thermocouple to insulate them electrically but not thermally.

A Lindeck potentiometer rather than a millivoltmeter is connected to the output emf of the converter to provide high resolution. A circuit diagram is shown in Fig. 3. The potentiometer has ranges of 0.15, 0.75, 7.5, and 15 mv. The two lowest ranges are used for sensitivity checks. With a galvanometer having a sensitivity of 200 mm (millimeter)/μa (microampere) it provides a selected resolution of either 2 or 10 μv (microvolts) per mm deflection. The potentiometer is completely shielded, and a 2-conductor shielded cable connects the output of the converter to the terminals marked *E* in the figure. The advantages of this potentiometer are its low thermal emf (less than 1 μv) and its freedom from drift (much less than 0.01%/minute under ordinary laboratory conditions).

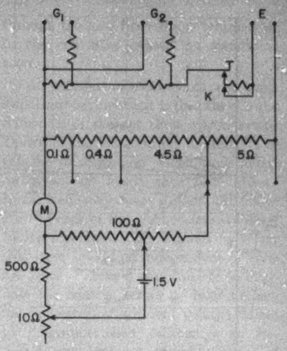


Fig. 3. Circuit of Lindeck potentiometer

M=1.5-ma d-c milliammeter
K=Thermofree key
E=Emf terminals, AN10SL
G_1, G_2=Galvanometer terminals, AN10SL
Note: All fixed resistors of manganin, accuracy +0.2%

Theory

Because of the geometry, the resistor, *R*, in the cylinder of each low-voltage converter (without inner shields) may be represented as a transmission line of length, *l*, with a uniformly distributed series impedance of *z* ohms per unit length and a uniformly distributed shunt ad-

Table I. Voltage Converters

Rated Voltage, Volts	Thermoelement Rated Current, Ma	Thermoelement Heater Resistance, Ohms	Series Resistor, Kilohms
1	5	90	0.2
3	5	90	0.5
5	5	90	1
10	5	90	2
20	10	25	2
50	10	25	5
100	10	25	10*
200	10	25	20*

* Two equal resistors in configuration of Fig. 2.

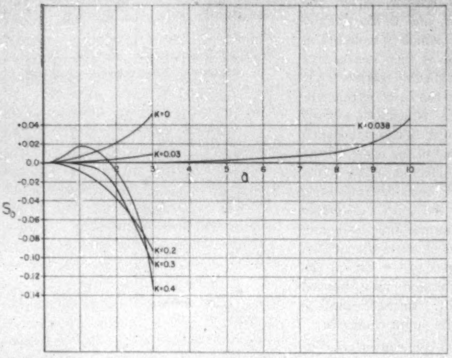


Fig. 4. Computed a-c–d-c difference as a function of the parameters a and k

mittance of y mhos per unit length. The line is terminated by the heater of the thermoelement. We are interested in the magnitude of the transimpedance $Z_c=V_t/I_o$ where V_t is the input voltage being measured and I_o is the current through the heater. If the heater resistance, R_h, is much less than that of the resistor, R, we may to a first approximation consider this a short-circuited line. It will be apparent later that this is not a drastic approximation at low frequencies since we are concerned only with the effect of the line reactance on Z_c. Then by ordinary steady-state transmission-line theory,

$$Z_c=\frac{Z}{\sqrt{ZY}}\sinh\sqrt{ZY} \qquad (1)$$

where $Z=z\ell=R+j\omega L$, and $Y=y\ell=j\omega C$ and R, L, and C are the total resistance, series inductance, and shunt capacitance of the line. We are interested in the low-frequency range in which the effects of the reactances are small.

We may define the parameters $a=\omega CR$ and $b=\omega L/R$. If each of these is less than unity the hyperbolic function may be approximated by the first three terms of its series expansion, so that

$$Z_c=Z\left(1+\frac{ZY}{6}+\frac{(ZY)^2}{120}+\ldots\right)$$

Then if all terms in a^nb^m for which $m+n>3$ are discarded we find that

$$Z_c\approx R(1+jb)\left(1-\frac{ab}{6}-\frac{a^2}{120}+\frac{ja}{6}\right)$$

and, to the same order,

$$|Z_c|\approx R\left(1+\frac{a^2}{90}+b^2-\frac{ab}{3}\right)^{1/2}\approx R\left(1+\frac{a^2}{180}+\frac{b^2}{2}-\frac{ab}{6}\right) \qquad (2)$$

The a-c–d-c difference of a voltage converter is defined as

$$S=\frac{V_{ac}-V_{dc}}{V_{dc}} \qquad (3)$$

where V_{ac} and V_{dc} are the rms alternating and direct voltages required to obtain the same response (output emf) of the converter. If the thermoelement has no a-c–d-c difference and if we define $k=b/a$, this becomes

$$S_o=\frac{|Z_c|-R}{R}\approx\frac{a^2}{2}\left(k^2-\frac{k}{3}+\frac{1}{90}\right) \qquad (4)$$

where S_o signifies that other possible causes of a-c–d-c differences are neglected.

The relation makes it possible to estimate S_o to this approximation very simply. For the cylindrical construction of Fig. 1 with a cylindrical unspiraled resistor, $L = 0.012\ell M$ microhenry and $C=0.61\ell/M$ picofarad, where $M=\log_{10}(g/h)$ and g and h are the diameters of the cylinder and resistor, respectively, and ℓ is in inches. Thus numerically, $k=1.9\times10^4(M/R)^2$ and $a=3.8\times10^{-12}f\ell R/M$, where f is the frequency.

For example if $R=2{,}000$ ohms, g and h are 2 in. (inches) and 1/8 in., and $\ell=2.0$ in., $S_o=+0.001$ (+0.1%) at 40 mc.

It is apparent from equation 4 that to this approximation certain values of k should make $S_o=0$. These are $k_1=0.038$ and $k_2=0.30$. The numerical values of resistance required are $R_1=710M$ and $R_2=250M$. Unfortunately the logarithmic relationship for M permits little range in R for reasonable values of g and h, for if g is greater than the length of the resistor, end effects may become pronounced.

It is of interest to extend the analysis to higher values of the parameter, a. This has been done by solving equation 1 for Z_c with the aid of Kennelley's tables and graphs of hyperbolic functions having complex arguments,[4] and then computing $S_o=(|Z_c|-R)/R$. The results are shown in Fig. 4. It is rather striking that the first solution of equation 4, $k_1=0.038$, results in reasonably small a-c–d-c differences (<5%) to $a=10$, a range much greater than that permitted by the initial assumptions.

For low-voltage converters the effect of the heater resistance R_h cannot be neglected. To the next approximation the heater can be considered as a lumped resistance termination for the transmission line. Then from steady-state transmission-line equations,

$$Z_c=R_h\cosh\sqrt{ZY}+\frac{Z}{\sqrt{ZY}}\sinh\sqrt{ZY} \qquad (5)$$

By expanding the hyperbolic functions and discarding higher order terms as before (for $a<1$ and $b<1$) we find that if $R_h<R$

$$S_o=\frac{|Z_c|-R_t}{R_t}\approx\frac{1}{2}\left(\frac{a}{1+m}\right)^2\left(k^2-Ak+\frac{B}{6}\right) \qquad (6)$$

where $R_t=R+R_h$, $R_h/R=m$, $A=m^2+2m/3+1/3$ and $B=m^2+2m/5+1/15$.

Table II. Comparative A-C–D-C Differences of Voltage Converters

Voltage Range			Comparative A-C–D-C Difference, D, Per Cent						
Converter A	Converter B	Applied Volts	0.1 mc	1 mc	5 mc	10 mc	20 mc	30 mc	40 mc
1	3	1	+0.01	0.00					−0.08
1	3	1.5		+0.01	−0.01	0.00	−0.02	−0.03	−0.06
3	5	2		0.00					−0.05
3	5	3	0.00	0.00	0.00	−0.01	−0.01	−0.03	−0.04
5	10	3		0.00					+0.14
5	10	5	0.00	0.00	−0.01	+0.01	+0.02	+0.07	+0.15
10	20	6		0.00					−0.02
10	20	10		0.00		0.00	0.00	−0.01	−0.02
20	50	15	0.00	0.00					−0.04
20	50	20				0.00	0.00	−0.01	−0.04
20	100	20	0.00	0.00	+0.02	+0.07	+0.24	+0.33*	+0.17
50	100	50		0.00				+0.33	
50	200	50	+0.01	0.00	−0.02	−0.06	−0.04	+0.19*	+0.57
100	200	90		0.00				−0.14	

Note: $D=100(S_B-S_A)$, where S is defined by equation 3, and the subscripts refer to converters A and B.

* Additional comparisons made at 25 and 35 mc.

The values of S_0 for the 1- to 20-volt converters have been calculated by equation 6 at a frequency of 40 mc, at which a and b are still less than unity. The results are given in the last column of Table II.

When $k<0.038$, distributed capacitance from the resistor to the cylinder predominates. At a given frequency and geometry, the resultant error is proportional to the square of the resistance. It can be minimized by the construction shown in Fig. 2 in which two inner cylindrical shields are used, with one connected to the input lead. Thus there are now two transmission lines in series. If $R_h<<R$, application of transmission line equations, with careful attention to signs, gives for the desired transimpedance,

$$Z_c=\sqrt{\frac{Z_1}{Y_1}}\tanh\sqrt{Z_1Y_1}\cosh\sqrt{Z_2Y_2}+\sqrt{\frac{Z_2}{Y_2}}\sinh\sqrt{Z_2Y_2} \quad (7)$$

where the subscripts 1 and 2 refer to the left and right hand lines (resistors and shields) of Fig. 2, and the resistance of the thermoelement is neglected.

If $Z_1=Z_2=Z$ and $Y_1=Y_2=Y$ we have

$$Z_c=\frac{2Z}{\sqrt{ZY}}\sinh\sqrt{ZY}$$

Thus the transimpedance is twice that of either line, and the a-c–d-c difference, S_0, will be that of half the total line. The parameter, a, of equation 4 will then be one-fourth that for a line of the same dimensions without inner shields. Thus at a given frequency for which $a<1$ the use of the double shield rather than a single cylinder of the same diameter should reduce the frequency error, for resistors of the same value and dimensions, by a factor of 16.

However, more detailed analysis indicates that even further improvement may be possible. For these converters $k<<0.038$, so that the series inductance of each line may be neglected. By expanding the hyperbolic functions of equation 7 and discarding higher order terms it can be shown by lengthy and rather tedious algebraic manipulation that $S_0\approx0$ when $\ell_1/\ell_2\approx1.032$, where ℓ_1 and ℓ_2 are the lengths of the left and right hand lines of Fig. 2, and $a<1$.

For the convenient dimensions shown in Fig. 1, the parameter, a, of each low-voltage (up through 20 volts) converter now being used is low enough that it has not been necessary to proportion the converters for minimum error in accordance with equation 6. The higher voltage elements are constructed with dual shields, with the grounded shield axially adjustable. This shield is set to give minimum frequency error as explained in the next section.

A useful theoretical study of coaxial lines, terminated by resistors, was published by Crosby and Pennypacker.[5] They show conditions for minimizing the input reactance.

Tests

Although long-time stability is not required of these converters, fluctuations or drifts in emf for the short time between the a-c use and the d-c calibration must be less than the desired accuracy. Such changes can arise from self-heating effects and ambient temperature changes, and from thermal emfs and other changes in the Lindeck potentiometer. Tests have shown that the self-heating of the carbon resistors, which have a load coefficient of about 1% per watt, is the largest source of drift. The change in resistance is very nearly exponential, with a time-constant of about 2 minutes. Since d-c calibrations can easily be made within 30 seconds of the a-c readings, this self-heating error is not significant in a-c tests if a short warm-up period is allowed. It is almost completely eliminated by the procedure used for a-c–d-c transfer tests. The effect of ambient temperature changes—the thermoelements have temperature coefficients up to 0.1%/C (degrees centigrade)—should also be insignificant in a laboratory with reasonable temperature control.

A great advantage of thermoelements for a-c–d-c transfer measurements is the almost complete electrical isolation of the input and output circuits. The uhf thermoelements have a small electrically insulating bead between the heater and hot junction of the thermocouple to eliminate conductive coupling. To minimize mutual inductance the plane of the thermocouple leads is at right angles to the heater and its colinear supports. Tests showed that at 4 mc the induced a-c voltage in the output circuit of each thermoelement was less than 20 μv, the resolution of the detector used. Induced currents in the thermocouple circuit can cause errors by joule heating of the thermocouple, but calculations indicate this should not be significant if the induced voltage is less than a few millivolts.

The shield of the Lindeck potentiometer provides reasonable immunity from induced fields. In some tests at 10 mc a current of 2 ma from the shield to ground through the lead from the potentiometer to the converter caused no significant error.

For transfer measurements the most important requirement is that the a-c–d-c difference of each of these converters be known to the full accuracy desired. The general principles on which such determinations are based have been given.[2] In the frequency range studied, the major error of each of these converters is caused by the reactance. For a given converter this error should be independent of voltage level. Thus, it was feasible to evaluate the relative errors by intercomparing converters of adjacent voltage ranges to determine their differences in frequency response. A complete series of such comparisons was made at two voltage levels for each pair of converters, at frequencies up to 40 mc. In each comparison the two converters were connected in parallel to a coaxial lead through a tee fitting (*GR 874*), and a shielded potentiometer was connected to each output. Each potentiometer was adjusted for zero deflection at the test voltage. The converters were then supplied in succession with alternating, direct, reversed-direct, and alternating voltage. Each voltage was adjusted to produce the same emf of the higher range converter and the deflection of the galvanometer connected to the other converter was observed. From the differences in emf, directly determined from these differences in deflection, the difference in the frequency response of the two converters was determined.[2] In these tests the movable inner shields of the 100- and 200-volt converters were adjusted for best performance over the desired frequency range, by comparison with the 20- and 50-volt converters. The errors of these adjustable converters were found to be complicated functions of the frequency and shield position.

The results of these intercomparisons from 0.1 to 40 mc are given in Table II. They show that for each pair of converters the relative a-c–d-c differences were independent of the applied voltage to 0.02% or less, and were less than 0.4% up to 30 mc. For the low-voltage converters they were less than 0.05% to 20 mc without exception and without evidence of systematic errors.

The relative a-c–d-c differences of most of the thermoelements used in the voltage converters were also determined before the thermoelements were installed, by making similar a-c–d-c intercomparisons as current-measuring elements at 40 mc. For these tests the two thermoelements were connected in series along the axis of a brass cylinder having the same dimen-

Table III. Results of Test of 1-Volt Voltmeter

Frequency, Mc	Applied Voltage,* Volts
1	0.997
10	0.998
30	0.998
100	0.999
200	0.982
300	0.945
400	0.883

*For same reading of millivoltmeter at each frequency

sions as in Fig. 1 but with a coaxial input connector at each end and with two 2-pin output connectors. A split cylinder with one part readily removable facilitated changing thermoelements. The input was applied to one connector with the other short-circuited. Rough calculations indicated that at 40 mc the current should change by about 0.5% along the transmission line formed by the heaters of the thermoelements in this cylinder, but that the distribution should not be greatly dependent on the resistance of the heaters, less than 100 ohms each. Therefore, two determinations of relative a-c–d-c differences were made with each pair of thermoelements, with first one end then the other of the cylinder short-circuited. For each pair the two determinations differed by about 0.4% at 40 mc, but in each case their algebraic average was less than 0.02%.

These calculations and tests also provide assurance that even at 40 mc the current along the heater of a single thermoelement terminating a voltage converter is well within 0.1% of the value at the mid-point of the heater.

The 1-volt converter was compared at rated voltage with a wire-wound thermal voltage converter of known a-c–d-c difference at 3 cps and 20 kc, with observed differences of +0.02% and 0.00% respectively.

The response of the 1-volt converter, with a shielded d-c millivoltmeter connected to its output, was determined to better than 1% at frequencies from 1 to 400 mc by the bolometer bridge of Selby and Behrent,[6] with the results shown in Table III. It is evident that the frequency influence is very small to 100 mc. It increases rapidly at higher frequencies, becoming −5.2% at 300 mc. The calculated a-c–d-c difference at 300 mc by equation 6 is only −0.9%; this converter has a 1/4-in.-diameter unsealed resistor. Some of the discrepancy may be accounted for by the voltage rise in the connector. A type *N-UG58/U* input connector was substituted for the type *874* for these tests.

As a check on equation 6 at larger a-c–d-c differences, the 50-volt converter was tested before its inner shield was installed. The measured a-c–d-c difference at 40 mc was +1.3% and was accurately proportional to the square of the frequency. The computed values were 40% smaller, indicating that the effective length of the resistor, the end effect, was 1.2 times the actual length. At 40 mc the measured a-c–d-c difference was changed by less than 0.02% when the resistor was mounted 3/32 inch off the axis of the cylinder, indicating that exact centering is not critical even when the errors are large.

The intercomparisons of Table II show that the a-c–d-c differences of the five low-voltage converters without inner shields, and with 1- to 20-volt ranges, are all equal to better than 0.05% to more than 20 mc. At 40 mc they agree to 0.1% with the values calculated by equation 6. Values cannot readily be computed for the higher voltage elements with the inner shields. This unanimity between converters having such a wide range of resistors and different thermoelements gives good assurance against unknown sources of error. It is quite unlikely, but not impossible, that each converter would have the same a-c–d-c error. However the test of the 1-volt converter with the bolometer bridge provides most valuable additional assurance. The large errors of this converter above 100 mc decrease rapidly with decreasing frequency, well within the stated accuracy of the bridge measurements. For converters without inner shields almost all known causes of such errors, such as the effect of reactance, equation 6, skin effects in the resistor or thermoelement, etc., should cause a-c–d-c differences approximately proportional to the square of the frequency, over the range for which the errors are small. Thus the authors believe that the large measured errors above 100 mc can be extrapolated downward to lower frequencies to indicate with considerable confidence that the a-c–d-c difference of this converter is less than 0.2% at 40 mc and less than 0.1% below 30 mc.

Based on the foregoing consideration the authors assigned a value of zero to the a-c–d-c difference of the 1-volt converter to 40 mc, and then determined the a-c–d-c differences of all the other converters from the intercomparison data in Table II. The results, rounded to the nearest 0.05% to 10 mc and 0.1% to 40 mc, are given as observed values in Table IV.

Use

These converters are most conveniently used to measure the a-c–d-c difference or frequency influence of other rms instruments such as thermocouple voltmeters, which are now available with accuracies of 1/2% to 10 mc. A typical setup for this is shown in Fig. 5. The scale calibration of the voltmeter can then easily be checked on reversed direct current. The a-c–d-c difference tests are similar to the intercomparisons already described, and d-c calibration of the converter is not necessary once the scale factor of the Lindeck potentiometer is determined for each converter. This is the per-cent change in input voltage per centimeter change of galvanometer deflection.[3] Either direct or low-frequency alternating current may be used as the reference frequency and the test may be made rapidly and accurately. The results are only slightly affected by drifts in either instrument. At frequencies above about 20 mc a small lead correction may be necessary if the connectors between the junction plane and the two instruments are not electrically equal, but this is readily determined to the required accuracy.

For a-c measurements a d-c potentiometer of 0.1% accuracy or better is required to measure the d-c reference voltage. A deflection potentiometer and volt box, or an automatic self-balancing potentiometer ("digital voltmeter") should be convenient for this. For testing electronic voltmeters frequency-response measurements (differences in reading for

Table IV. A-C–D-C Differences of Voltage Converters

Rated Voltage, Volts	Per-Cent A-C–D-C Differences, 100S					
	Observed					Calculated
	to 5 mc	10 mc	20 mc	30 mc	40 mc	40 mc
1	<0.05	<0.05	0.0	0.0	0.0	0.0
3	<0.05	<0.05	0.0	0.0	−0.1	0.0
5	<0.05	<0.05	0.0	−0.1	−0.1	0.0
10	<0.05	<0.05	0.0	0.0	0.0	+0.1
20	<0.05	<0.05	0.0	0.0	0.0	+0.1
50	<0.05	<0.05	0.0	0.0	0.0	
100	<0.05	+0.05	+0.2	+0.3	+0.1	
200	<0.05	−0.05	0.0	+0.2	+0.5	

Fig. 5. 5-volt converter setup for a-c–d-c test of experimental thermocouple voltmeter. Lindeck potentiometer at right, double-shield converter and experimental split-shield element in foreground, a-c–d-c switch not shown

the same voltage at the test and reference frequencies) should be particularly convenient. The form factor and crest factor of the a-c sources must be within 0.1% of the values for a sine wave, 1.111 and 1.414 respectively, since most electronic voltmeters respond essentially to the average or crest (peak) values. Since these factors depend upon the phase angle as well as the magnitude of each harmonic it would ordinarily be necessary to make sure that the ratio of the magnitude of the nth harmonic to the fundamental does not exceed $0.1xn\%$ when an average-reading instrument is tested and 0.1% when a crest-reading instrument is tested. The results should then differ from those obtained with a sine wave by less than 0.1%.

An r-f generator of at least 5 watts output and good voltage stability is required but only moderate frequency accuracy and stability are needed. It is very difficult to construct broad-band generators of good wave form at this power level. However, a simple adjustable L/C (inductance capacitance) tuned circuit, with the instruments connected across the capacitor, can be used simultaneously to improve the wave form, match the impedance of the instruments to that of usual 50-ohm source, eliminate the capacitance loading of the instruments on the source, and provide the higher voltages often required. For these combined purposes the values of inductance and capacitance should be chosen so that $\omega L = 1/\omega C \approx \sqrt{R_1 R_2}$, where R_1 and R_2 are respectively the resistances of the source, including the inductor, and the load (instruments). For low voltages a 50-ohm resistor can be connected in series with the L/C circuit and the instruments and attenuators as needed can be connected in parallel with this resistor for improved wave form.

Conclusions

Cylindrical film resistors in series with a thermoelement in a coaxial line makes possible single-range thermal voltage converters that are useful as a-c–d-c transfer instruments over a very wide frequency range, 3 cps to at least 30 mc, with unusually high accuracy. They are inexpensive and easy to construct. Their frequency influence may be estimated by reasonably simple equations, with results which agree well with the measured values up to 40 mc. The voltage converters may be useful to considerably higher frequencies. The higher range converters require considerable power, up to 2 watts, and have a marked but short warm-up drift. These disadvantages could be reduced by using metal-film resistors of higher resistance with a 5-ma thermoelement, but this would very probably reduce the frequency range for the same attainable accuracy. The good performance of these converters was obtained with deposited-carbon resistors of a new type. Earlier converters with resistors of different construction showed discrepancies up to 0.5% at 40 mc. Further work is planned to determine the cause of these differences.

These voltage converters can be used quickly and easily to make a-c–d-c difference tests to determine the frequency influence of other rms instruments to 0.1% or better to at least 10 mc and to 0.2% at 30 mc. Direct a-c measurements are made by the transfer technique, which in most cases can be arranged for reasonable simplicity. With sources of suitable waveform, average-reading and crest reading instruments could also be calibrated if desired. In all of these applications the applied frequency need not be closely determined or held because of the flat frequency response of these voltage converters.

References

1. ELECTRICAL INDICATING INSTRUMENTS. *ASA C39.1-1959*, American Standards Association, Inc., New York, N. Y., sect. 3.11.3, May 1959.

2. THERMAL CONVERTERS AS AC-DC TRANSFER STANDARDS FOR CURRENT AND VOLTAGE MEASUREMENTS AT AUDIO FREQUENCIES, F. L. Hermach. *Journal of Research*, National Bureau of Standards, Washington, D. C., vol. 48, 1952, pp. 121–38.

3. A WIDE-RANGE VOLT-AMPERE CONVERTER FOR CURRENT AND VOLTAGE MEASUREMENTS, F. L. Hermach, E. S. Williams. *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 78, Sept. 1959, pp. 384–88.

4. TABLES AND CHART ATLAS OF COMPLEX HYPERBOLIC AND CIRCULAR FUNCTIONS, A. E. Kennelly. Harvard University Press, Cambridge, Mass., 1921.

5. RADIO-FREQUENCY RESISTORS AS UNIFORM TRANSMISSION LINES, D. R. Crosby, C. H. Pennypacker. *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 34, 1946, pp. 62–66.

6. A BOLOMETER BRIDGE FOR STANDARDIZING RADIO-FREQUENCY VOLTMETERS, M. C. Selby, L. F. Behrent. *Journal of Research*, National Bureau of Standards, vol. 44, 1950, pp. 15–30.

◆

Discussion

Morley J. Lush (Rawson Electrical Instrument Co., Cambridge, Mass.): I feel that the instruments described in this paper are a real contribution to the art of accurate voltage measurements at high frequencies. Although it is difficult to make resistors with low reactances, Mr. Hermach has shown that these can be obtained, and that voltmeters can be constructed which have a flat frequency response all the way from direct current out to the megacycle region. The frequency response will remain stable for long periods of time, so only d-c calibrations are needed, once the frequency response has been determined. I have found from my own work that accuracies of 1% can be obtained at frequencies up to 100 mc and now these instruments are available on the market. A precision of 0.1% can be obtained, but the absolute accuracy is limited by the lack of a primary standard. I have been told that the National Bureau of Standards expects to be able to offer 0.25% certifications by the end of 1960.

Other workers in this field have used a fixed or variable capacitor as the attenuator element in series with the thermocouple. It is possible to make an excellent low-loss capacitor using a micrometer screw for adjustment. A wide frequency range and wide voltage range can be obtained on one instrument. However, the voltmeter cannot then be used for d-c to a-c transfer measurements, since it has zero response at zero frequency. Indeed, the response is directly proportional to the frequency, and it is necessary to calibrate the voltmeter at each voltage and frequency for which it will be used.

It should be pointed out that connections to these voltmeters are a problem, as they are in all high-frequency apparatus. The use of coaxial connectors can be misleading, as any appreciable length of 50-ohm coaxial line will cause errors due to standing waves, unless the voltmeter input impedance is made exactly equal to the impedance of the line. Careful matching of the line impedance will certainly be necessary if voltmeters of this type are to be used at frequencies above 100 mc.

F. L. Hermach and **E. S. Williams:** Mr. Lush's remarks are appreciated. We believe these thermal voltage converters, by taking advantage of the powerful a-c–d-c transfer technique, should make possible a new order of accuracy in a-c voltage measurements up to at least 30 mc. Their usefulness will probably fall off rapidly at higher frequencies, except for the lower voltage ranges, up to perhaps 10 volts.

A major advantage of an a-c–d-c transfer standard is that its a-c–d-c difference is relatively permanent so that once evaluated, it should not ordinarily need to be redetermined. Thus, elaborate repeated a-c calibrations are not required. This and the ease and simplicity of constructing these converters should offset the extra labor of making the a-c to d-c transfer for each measurement.

We mentioned connection errors only briefly. Each converter is a high-impedance load at the end of a short coaxial line from the junction plane to the converter. If this line is a small fraction of a wavelength the voltage rise from the junction to the end of the line is approximately $50h^2\%$, where $h=\omega CZ_0$ and C and Z_0 are the total shunt capacitance and the characteristic impedance of the line. At 30 mc for a 2-inch length of typical 50-ohm line the correction is only 0.1%. It is offset by a similar correction if an equal line is used between the junction plane and the test instrument or load to be measured. At higher frequencies or for longer lines the correction can be large and this approximate formula may no longer be valid. In such cases, however, voltage measurements probably have meaning only with lines terminated in their characteristic impedance, so that the standing-wave ratio is close to unity.

D-C Magnetic Crane Hoist Control for A-C Powered Cranes

A. H. MYLES
FELLOW AIEE

M. C. DAVIES
ASSOCIATE MEMBER AIEE

L. J. SRNKA
NONMEMBER AIEE

Synopsis: The use of a d-c series-wound motor, series-wound motor brake, and power-type overhoist limit switch for the hoist drive of an otherwise all a-c powered crane is highly desirable. Time proven safety and simplicity of d-c dynamic lowering hoist systems can now be further improved in performance, and better motor stability is possible when d-c power is made available through metallic plate rectifiers.

Methods of calculating motor performance under dynamic lowering connections are discussed.

FIFTY YEARS of development of crane hoist control systems have transpired since the first dynamic lowering control method was introduced to eliminate the need for mechanical load brakes. Moreover, during this time probably no other industrial control application has received more attention than has crane hoist control for both d-c and a-c motors and their respective power supply. It should be mentioned in this connection that users, crane builders, and control manufacturers have all generally favored the crane that is operated by d-c series motors because it offers advantages such as simplicity, low maintenance, and maximum safety from the use of series-wound brakes and power-type limit switches. On the other hand, crane control for a-c wound rotor motors has also been greatly improved and, in some designs, gives performance characteristics matching those obtained with direct current.

Modern power distribution systems in industry are generally 3-phase and, through properly connected transformers, provide a-c service at 230, 460, or 550 volts. Also, with the exception of the steel industry, large capacity d-c conversion equipment is usually not available for d-c hoist motors on cranes. This means, therefore, that, lacking a d-c source of supply for a crane hoist, it is desirable to make use of the recently improved metallic type of rectifiers, individually mounted for each application; such installations occupy a comparatively small space, are reasonably low in cost, and can be recommended to deliver dependable d-c power.

When a nominal 230-volt 3-phase source is used with a set of six bridge-connected rectifiers, the d-c output under loaded conditions is approximately 300 volts or 1.3 times the a-c input. This implies that a 230-volt d-c series motor will develop 1.3 times its 230-volt horsepower rating when used in such a rectifier-conversion system since for a given load current and developed torque, the motor speed is directly proportional to the voltage impressed across the armature terminals. Moreover, the use of the higher 300-volt source will have no damaging effect upon the motor since mill-type motors are designed to tolerate the higher voltage and are, in fact, regularly used in adjustable voltage systems at the higher potentials.

When the a-c system is supplied at 460 volts, it is recommended that a transformer be used to step down the input voltage to the rectifier to 230 volts. Three-phase rectification of 460 volts a-c will provide approximately 600 volts d-c. If a 550 volt d-c motor were considered, the increased speed and horsepower would be only 9% and the disadvantages associated with using the higher power and control circuit voltage as well as the higher motor cost could not be justified.

When a comparison is made between a d-c series motor and an a-c wound rotor motor, consideration must be given to each portion of the hoisting and lowering functions. An a-c crane hoist controller using a wound rotor motor will hoist or lower all normal loads at a maximum speed of near synchronous speed. In this connection, it should, however, be understood that a normal crane-hoist operates in the following cycle: (1) lowering an empty hook; (2) hoisting full load or less; (3) lowering full load or less; and (4) hoisting an empty hook. Motor horsepower calculations indicate that when used at 300 volts d-c the motor should have a normal 230-volt rating that is approximately 50% to 55% of its a-c counterpart, assuming that each one will complete its operating cycle in the same time. This advantage, possessed by the d-c motor, results from increased speed at the higher than normal voltage, the inherent higher light load speeds hoisting of the series motor, and the higher light load and heavy load lowering speeds of the d-c dynamic lowering control.

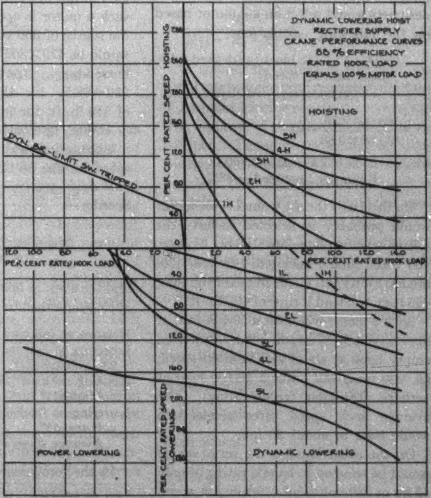


Fig. 1. Crane performance curves for d-c dynamic lowering control with rectified power supply

---

Paper 60-30, recommended by the AIEE Industrial Control Committee of the Industry Division and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., January 31–February 5, 1960. Manuscript submitted March 19, 1959; made available for printing November 18, 1959.

A. H. MYLES, M. C. DAVIES, and L. J. SRNKA are with the Square D Company, Cleveland, Ohio.

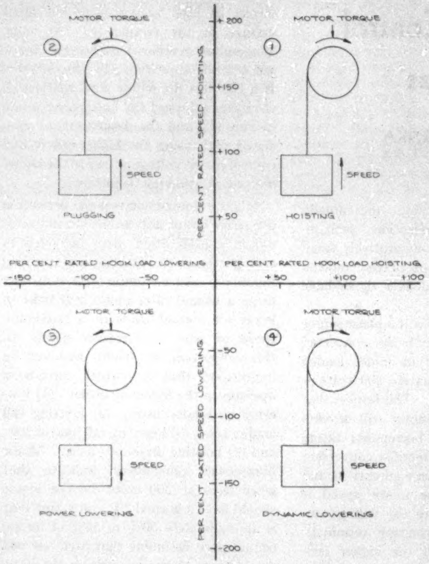


Fig. 2. Four-quadrant representation of crane hoist performance

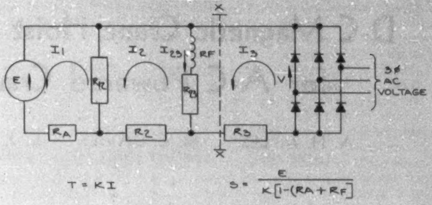


Fig. 3. Equivalent mesh network circuit for dynamic lowering hoist control with rectifier power supply

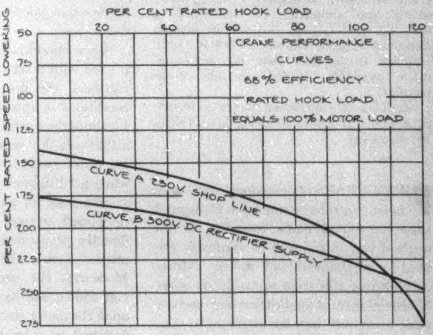


Fig. 4. Stability performance curves at maximum lowering speeds with 230-volt normal supply and with 300-volt rectifier power supply

To illustrate the foregoing discussion, consider a 50-ton crane hoist to give a full-load hoisting speed of 25 feet per minute (fpm), assuming a 460-volt 3-phase wound rotor motor. The crane designers would select an a-c motor based on the following calculation:

$$\text{hp} = \frac{(\text{payload} + \text{hook block}) \times (\text{desired full-load hoisting speed})}{33{,}000 \times \text{crane efficiency}}$$

$$= \frac{103{,}000 \times 25}{33{,}000 \times 0.88} = 89$$

where: hook block weighs 3,000 lb.
Efficiency (assumed) = 88%

On this basis the a-c wound rotor motor would probably be selected to have the following nameplate data: 100 hp, 3 phase, 460 volts, totally enclosed, 1/2 hour (30-minute) duty, 75 C (degrees centigrade), 865 rpm, full load torque of 600 lb.-ft.

Such a motor will hoist a 50-ton load at approximately 870 rpm, hoist or lower an empty hook at about synchronous speed, i.e. 900 rpm, and lower a 50-ton load at perhaps 920 rpm (regenerating). The average hook speed after acceleration would then be about 26 fpm.

On the other hand, a d-c series motor would be selected to have a rating of only 45 hp at 230 volts and 470 rpm; it would be totally enclosed with a 1/2 hour (30-minute) duty at 75 C and develop a full-load torque of 510 lb.-ft. However, when such a motor is operated from a rectified d-c source of 300 volts, its full load speed would be 470×300/230 = 610 rpm and its horsepower output would be 45×300/230 = 58.5; for the same load torque of 510 lb.-ft. the full load current would, of course, be equal to that at 230 volts.

Calculating next the full load hoisting speed for the 50-ton crane,

$$\text{Speed} = \frac{\text{hp} \times 33{,}000 \times \text{efficiency}}{103{,}000}$$

$$= \frac{58.5 \times 33{,}000 \times 0.88}{103{,}000} = 16.5 \text{ fpm}$$

and the average hook speed for a complete cycle would then be determined as follows:

Hoisting at full load	16.5 fpm
Lowering at full load (235% of 230 volt speed)*	29.8 fpm
Hoisting no load (250% of 230 volt speed)*	31.7 fpm
Lowering no load (175% of 230 volt speed)*	22.2 fpm

Average speed = 100.2/4 = 25 fpm

In the foregoing calculations it should be noted that the speeds are based on the values at no load and full load. In practice, however, the average load would probably be less than the full-load rating and this would therefore give an average speed that is somewhat higher than that obtained above. It should also be observed that the cycle of operations will be completed in the required average time although the hoisting speed of 16.5 fpm is somewhat less than the 25-fpm value.

In discussing dynamic lowering hoist control as applied to the d-c series motor, industry has established the conventional method of representing motor operation in four quadrants; this is shown in Fig. 2. Quadrant 1 is for power hoisting; quadrant 2 is for plugging from a hoisting to a lowering direction; quadrant 3 is for power lowering; and quadrant 4 is for dynamic lowering of overhauling loads. When analyzing crane hoist performance curves for a given motor in the four quadrants indicated, it should be pointed out that due to loading and hoist efficiency the actual speed regulation with load is always better than indicated by motor characteristic curves. This means that

---

* See curves, Fig. 1.

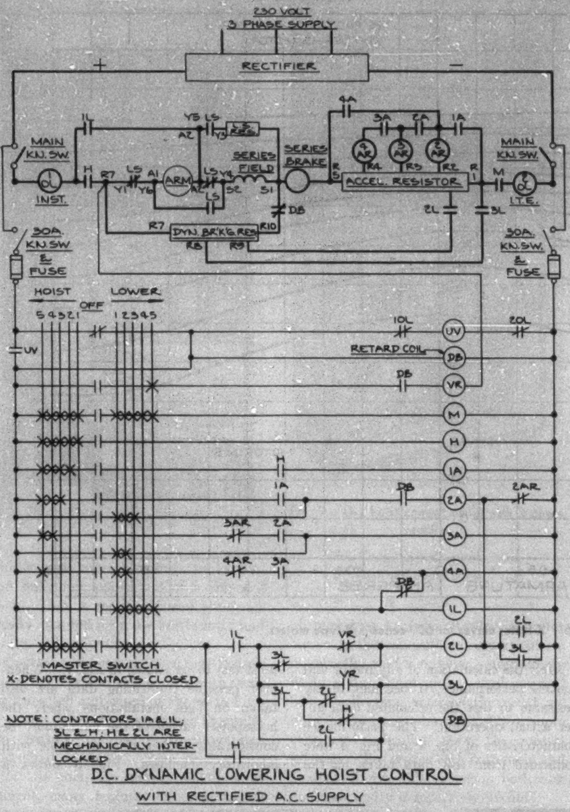

230 VOLT
3 PHASE SUPPLY
RECTIFIER
+
−
MAIN KN. SW.
INST.
OL
1L
H
R7
LS
A1
Y1
Y6
ARM
A2
Y5
Y3
L.S. RES.
LS
Y4
SERIES FIELD
S2
S1
SERIES BRAKE
R5
4A
3A
2A
1A
4AR
3AR
2AR
R4
R3
R2
R1
ACCEL. RESISTOR
M
MAIN KN. SW.
L.T.E.
DB
2L
3L
R7
DYN. BR'K'G. RES.
R10
R8
R9
30A. KN. SW. & FUSE
30A. KN. SW. & FUSE
HOIST
LOWER
OFF
5 4 3 2 1
1 2 3 4 5
UV
1OL
2OL
UV
RETARD COIL
DB
VR
M
H
1A
2A
3A
4A
1L
2L
3L
DB
MASTER SWITCH
X-DENOTES CONTACTS CLOSED
NOTE: CONTACTORS 1A & 1L; 3L & H; H & 2L ARE MECHANICALLY INTERLOCKED
D.C. DYNAMIC LOWERING HOIST CONTROL
WITH RECTIFIED A.C. SUPPLY

Fig. 5. Elementary wiring diagram of d-c dynamic lowering crane hoist control

motor loading and hoist efficiency should be carefully defined if the performance curves are to be properly interpreted.

With the high efficiency of modern crane hoists, the tendency of overspeeding an empty hook as it moves upward must be given consideration. This is because the stopping distance when entering the power limit switch may be excessive and commutator burning may be severe under conditions of strong dynamic braking. It is now common practice to apply a high resistance permanent armature shunt to such applications to limit the no load hoisting speed to 250% of rated value. With the application of 300-volt rectified d-c to the hoist motor, the permanent armature shunt resistor becomes a standard feature.

In the lowering of overhauling loads with d-c dynamic braking control, as represented by the fourth quadrant of Fig. 2, one feature of the control is that it is capable of providing high-speed light-load lowering while maintaining satisfactory motor stability when lowering heavy loads. With the usual constant potential dynamic lowering control scheme, the maximum no-load-lowering speed must be limited to approximately 140% of rated value since with high armature current any further field weakening would cause sufficient armature reaction to affect motor stability adversely. However, when d-c power is derived from a rectified source, the circuit constants are quite different. The equivalent mesh circuit for the dynamic lowering hoist control with a rectifier power supply is shown in Fig. 3. The performance curves of Fig. 4 compare the results obtained when lowering loads at maximum speed with power from a 230-volt shop line and with a rectified 300-volt supply. Calculations for the data for the curves are on a per-unit basis. Note particularly the improved stability and the higher no load speed of the rectified d-c operation with respect to the shop line operation. The elementary wiring diagram of the complete control is illustrated in Fig. 5.

Referring to Fig. 3, calculations for a 3-mesh circuit involve the following equations:

$$R_{11}=R_A+R_{12} \tag{1}$$

$$R_{22}=R_{12}+R_2+R_{23}+R_F \tag{2}$$

$$R_{33}=R_F+R_{23}+R_3 \tag{3}$$

Solving for the three loop currents, using a method similar to that given in some detail in the succeeding analysis for a 2-mesh circuit,

$$I_1=1/D[(R_{12}R_{23}V)-(R_{22}R_{33}-R_{23}^2)E] \tag{4}$$

$$I_2=1/D[(R_{23}R_{11}V)-(R_{33}R_{12}E)] \tag{5}$$

$$I_3=1/D[(R_{11}R_{22}-R_{12}^2)V-(R_{12}R_{23}E)] \tag{6}$$

where

$$D=(R_{11}R_{22}R_{33})-(R_{11}R_{23}^2+R_{33}R_{12}^2) \tag{7}$$

$$I_{23}=I_3-I_2 \tag{8}$$

$$E=\frac{1/D(R_{23}R_{11}V)-1/D(R_{11}R_{22}-R_{12}^2)V+I_{23}}{1/D(R_{33}R_{12}-R_{12}R_{23})} \tag{9}$$

Finally, using values of K from the curves of Fig. 6,

$$T=KI_1 \tag{10}$$

$$S=\frac{E}{K[1-(R_A+R_F)]} \tag{11}$$

In solving for values of torque T and speed S it becomes apparent that the value of I_3 cannot become negative, i.e. regenerative, because of the blocking effect of the rectifier. Therefore, in solving for overhauling load values of T and S it should be clear that the equivalent circuit to the left of X-X, Fig. 3, should be used; this is equivalent to saying that the circuit to the right of X-X is open-circuited ($I_3=0$) and that the developed power for the 2-mesh circuit is provided by the overhauling load.

Referring to the 2-mesh circuit to the left of X-X in Fig. 3, the following equations may be written:

$$E+I_1R_{11}-I_2R_{12}=0 \tag{12}$$

$$-I_1R_{12}+I_2R_{22}=0 \tag{13}$$

from which

$$I_1=-E\left[\frac{R_{22}}{R_{11}R_{22}-R_{12}^2}\right] \tag{14}$$

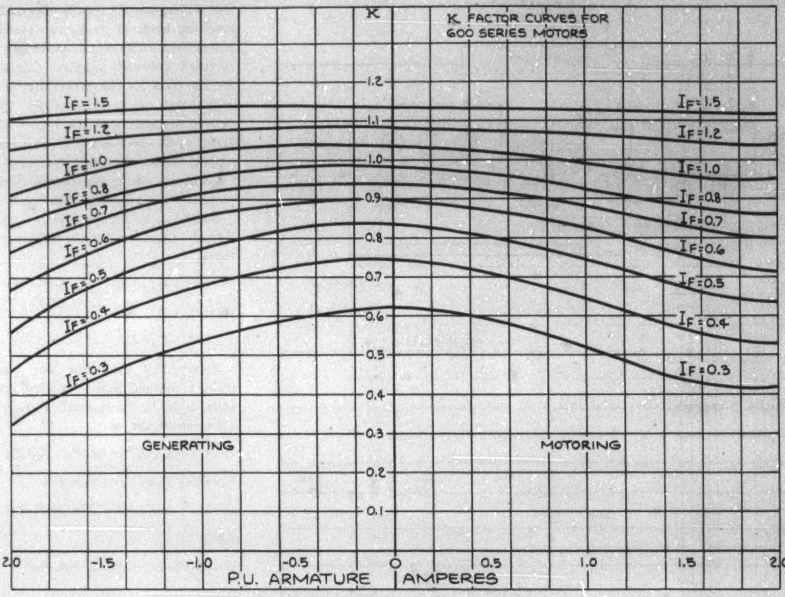


Fig. 6. K factor curves for 600-series mill-type motors

$$I_2 = -E\left[\frac{R_{12}}{R_{11}R_{22}-R_{12}^2}\right] \qquad (15)$$

The torque T and speed S values, equations 10 and 11, can then be determined as for the 3-mesh circuit.

After the calculation of any motor and control performance, it becomes highly desirable to test the calculated data under actual operation. The performance characteristics of Fig. 1 and Fig. 4 were confirmed from test data taken on the hoist test tower shown in Fig. 7. Whenever possible confirming data are also taken on field installations where the horsepower rating of the motor may be considerably higher than available with laboratory facilities. Fig. 8 shows a

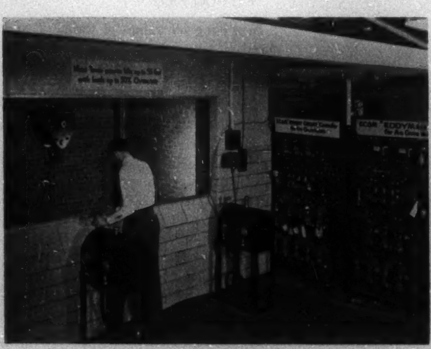


Fig. 7 (above). Experimental test hoist tower providing facilities for varied a-c and d-c circuit testing

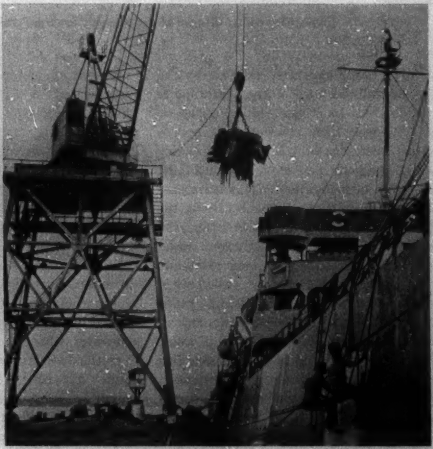


Fig. 8 (right). Whirler-type dock crane for scrap and general cargo handling. Power supply is a-c with rectifier d-c for hoist motor and control

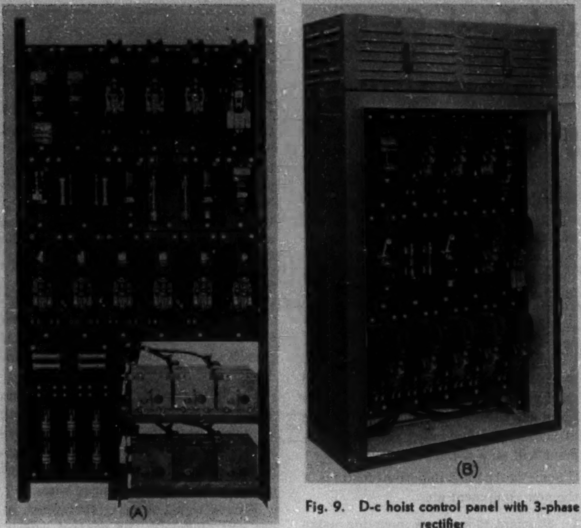


Fig. 9. D-c hoist control panel with 3-phase rectifier

A—Open panel construction
B—Enclosed panel construction

whirler-type crane handling the loading and unloading operations of a shipping dock. Here a rectifier power supply is ideally adapted for a d-c hoist motor and its control equipment.

Control panel construction may feature either open or enclosed control with rectifier mounted and wired as shown in Figs. 9(A) and (B) or the control may be similar but feature completely front-mounted and wired devices.

In the complete installation of crane control, many features affect mounting space, structure, weight, wiring, etc., and one costly and often difficult problem of mounting concerns trolley collector bars. In this connection it is particularly noteworthy that a rectifier power supply using a d-c motor and control requires but 4 collectors, whereas the usual a-c hoist control must have 10, 12, or more collectors depending upon the control schemes. For cranes with both main and auxiliary hoists this saving in both cost and complexity is indeed significant.

Finally, in any motor installation there is always the problem of line drop due to load current. Here again the rectified d-c supply has the advantage that the IR drop is a smaller per cent of the nominal 300-volt d-c system than is the same drop with the lower 230-volt d-c circuit thus giving improved load-speed regulation.

Summary and Conclusions

When a set of bridge-connected metallic rectifiers is used to supply power to a d-c series motor and its crane hoist control, the d-c output voltage will be about 1.3 times the line-to-line a-c 3-phase input voltage. This means that for a 230-volt a-c supply, a 230-volt d-c series motor will be energized from 300 volts, with a consequent increase in speed and horsepower ratings. In addition to the improvement in motor performance indicated, the following further advantages may be listed:

1. Faster hoisting and lowering speeds are obtainable.
2. Greater stability results when lowering heavy loads than is possible with a normal 230-volt d-c supply.
3. The d-c series motor develops better hoisting characteristics and faster acceleration than does the a-c wound rotor motor.
4. The d-c control scheme involving a series-wound motor, a series-wound brake, and a power-type overhoist limit switch provides maximum in safety and simplicity.
5. The d-c control scheme requires only 4 collectors, whereas the usual a-c hoist control must have 10 or more collectors.
6. Voltage regulation due to motor and lead drop is better on the 300-volt d-c system than with a normal 230-volt d-c shop line.

An All-Transistorized Trunk Carrier System

F. H. GARDNER
NONMEMBER AIEE

THE RAPID expansion of suburban and small urban communities has created such a demand for increased communication facilities that the telephone companies are hard pressed to supply the requirements. They find themselves required to supply not only house phones, business phones, etc., but new subscriber lines, new extended area trunks, and new toll trunks in order to handle the increased load.

Telephone companies are service organizations which are in business for the purpose of serving their customers quickly, efficiently, economically, and for making a profit for their stockholders. It is, therefore, necessary to provide added facilities which are reliable, quick and easy to install, simple to maintain, economical to operate, and require only modest investment in equipment.

Among the most expensive additions to a telephone company's plant are the new subscriber lines extended area service (EAS) trunk lines, and toll lines required to handle the increased traffic. The high cost of adding physical circuits may make it necessary to provide service by other means.

Discussion

The installation of subscriber carrier and trunk carrier equipment is often the most economical means of adding several lines since extensive plant wiring changes or additions are not required. As an answer to this problem Stromberg-Carlson has recently announced their All-Transistorized *565* Trunk carrier system as a companion to their *561* Subscriber Carrier system which has been on the market for several years. The *565* Trunk Carrier equipment is presently designed to provide five carrier channels for communication and *E&M* signaling. At this writing, the expansion of the system to a greater number of channels is contemplated.

DESCRIPTION

The equipment is of the amplitude-modulated double-sideband, transmitted carrier type. Fig. 1 shows a complete terminal. Those acquainted with the Stromberg *561* Subscriber Carrier Equipment will note the similarity.

Some of the most important features of the new equipment are:

1. Standard 19-inch rack mounting makes installation easy.
2. Plug-in major assemblies simplify service problems.
3. Test jacks in d-c, voice, and signaling circuits speed repairs.
4. Requires only 70 ma (milliamperes) per channel at −48 volts providing operating economy.
5. Operation from standard central office battery provides reliability.
6. Transmits over 150- or 600-ohm lines aleviating extensive plant changes.
7. Optional voice-frequency (v-f) termination increases utility, 600 ohms balanced 2 wire, 900 ohms balanced 2 wire, 600 ohms unbalanced to ground 4 wire are provided.
8. A per-cent make control makes 0% pulse distortion possible.
9. Wide bandwidth of 250 to 3,000 cps (cycles per second) extends area of application.

At the top of Fig. 1 is the line coil panel which mounts of v-f low-pass filter, a high-frequency (h-f) matching transformer for coupling to a 150- or 600-ohm circuit, and an inductance-capacitance (L-C) battery supply filter, and a 69-kc low-pass filter.

The lower five units are the five carrier channel units. These units each include 10 plug-in assembleis which are, from left to right:

1. Terminal card equipped with relay contact protection circuit, fuse, jacks to facilitate testing, etc.
2. Transmit band-pass filter card.
3. Receive band-pass filter card.
4. H-f card containing attenuating networks.
5. Transmitting card.
6. Crystal oscillator and hybrid coil and relay assembly.
7. Receiver card.
8. Signaling card containing filters, signaling oscillator, and signal amplifier.
9. V-f transmitting card containing attenuating networks (optionally a compressor may be used).
10. V-f receiving card containing attenuating networks (optionally an expandor may be used).

THE SYSTEM

Fig. 2 is a block diagram of a *565* Trunk Carrier system. The v-f, *E*, and *M* circuits are connected at the locations indicated.

The physical circuit is essentially no different from any other physical trunk circuit except that v-f low-pass filters have been inserted as shown to allow transmission of the additional five carrier channel signals on the same pair of wires.

The v-f signal entering the system at one end modulates the channel carrier. The *M* lead signal (−48 volts d-c) entering the system at *M* keys a 3,500-cps tone which in turn also modulates the channel carrier.

The modulated carrier of channels 3, 4, and 5 is fed directly to the line matching transformer. The carriers of channels 1 and 2, however, are fed to the line matching transformer through a 69-kc h-f low-pass filter which eliminates interference between low-frequency and high-frequency channels.

After the carrier signals arrive at the far end they are passed through trans-

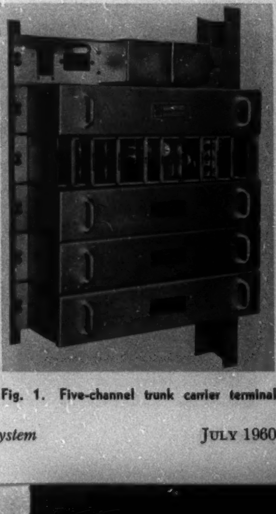


Fig. 1. Five-channel trunk carrier terminal

Paper 60-44, recommended by the AIEE Wire Communication Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., January 31–February 5, 1960. Manuscript submitted September 16, 1959; made available for printing January 27, 1960.

F. H. GARDNER is with Stromberg-Carlson Company, Rochester, N. Y.

The writer wishes to acknowledge the work of D. Jamieson of Stromberg-Carlson who did much of the circuit design, the work of Mr. P. May, Mr. T. Stump, and Mr. G. Sager, all of Stromberg-Carlson in special filter designs, and the co-operation of Mr. R. Eyler of Merchants and Farmer's Telephone Company of Montpelier, Vt., in offering the use of plant facilities for the field trial.

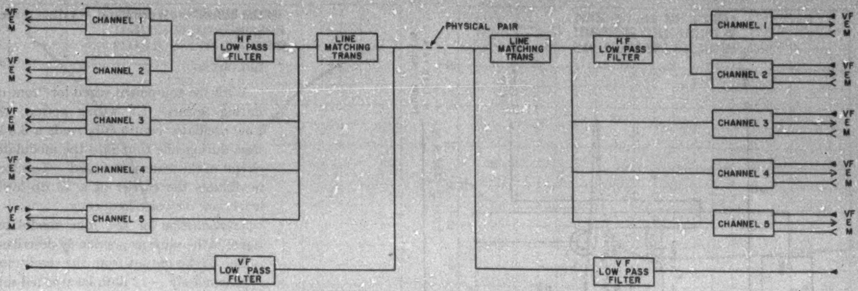

VF
E
M
CHANNEL 1
CHANNEL 2
CHANNEL 3
CHANNEL 4
CHANNEL 5
H F LOW PASS FILTER
LINE MATCHING TRANS
PHYSICAL PAIR
LINE MATCHING TRANS
H F LOW PASS FILTER
VF LOW PASS FILTER
VF LOW PASS FILTER

Fig. 2. Block diagram of 5-channel trunk carrier system

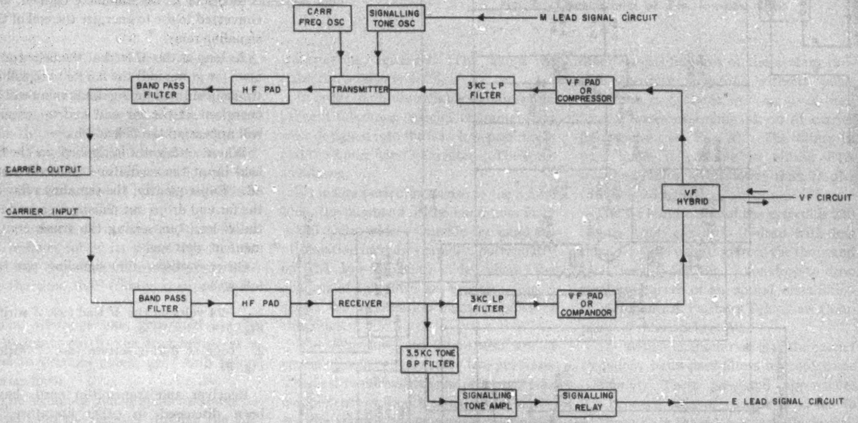

CARR FREQ OSC
SIGNALLING TONE OSC
M LEAD SIGNAL CIRCUIT
BAND PASS FILTER
H F PAD
TRANSMITTER
3KC L P FILTER
V F PAD OR COMPRESSOR
CARRIER OUTPUT
CARRIER INPUT
VF HYBRID
V F CIRCUIT
BAND PASS FILTER
HF PAD
RECEIVER
3KC LP FILTER
V F PAD OR COMPANDOR
3.5 KC TONE B P FILTER
SIGNALLING TONE AMPL
SIGNALLING RELAY
E LEAD SIGNAL CIRCUIT

Fig. 3. Block diagram of channel equipment

former and filter circuits duplicating those at the near end and are demodulated.

The voice signals may be heard at v-f. Signaling pulses having exactly the same % make as those delivered to the equipment on the near end *M* lead will appear at *E*.

THE TERMINAL

Fig. 3 is a block diagram of a single channel at one termnal.

VOICE TRANSMISSION

Looking at the voice transmission circuits, first note that the v-f signal enters at the v-f hybrid and travels to the transmitting v-f pad where the level is adjusted, thence through the 3-kc low-pass filter to the transmitter.

In the transmitter the voice-frequency signal modulates the carrier generated by the carrier oscillator which is transistorized and crystal controlled.

The modulated carrier amplified by two transistors in the transmitter card passes through the h-f pad where it may be attenuated as desired.

Next in its path is the band-pass filter which passes the carrier frequency and its sidebands only. The signal is thereby made available for transmission through additional filters and/or line matching transformers to a distant point over cable or wire lines.

Considering the lower half of Fig. 3 as the far-end receive section of the channel, the modulated carrier enters through a band-pass filter which rejects all other carrier frequencies. It then passes through the h-f pad to be attenuated as required to provide proper input level to the receiver.

The completely transistorized receiver amplifies, detects, and controls the output level within ±1 db (decibels) for carrier input levels varying between −13 and −28 dbm (decibels below 1 milliwatt).

From the receiver the voice signals are passed through a 3-kc low-pass filter which will stop any 3.5-kc signaling tones. In the receiving v-f pad the channel output may be adjusted to the desired level. Normal v-f input and output levels at

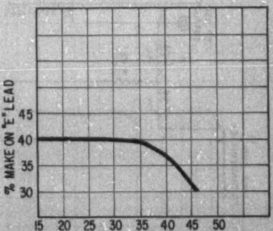


Fig. 4. Line loss in db at carrier frequency

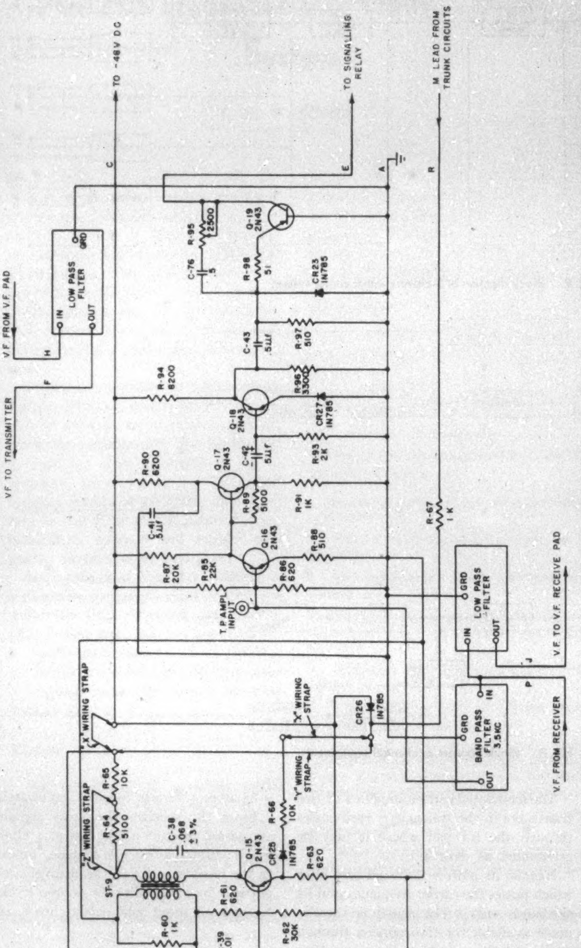

TO -48V DC
TO SIGNALLING RELAY
M LEAD FROM TRUNK CIRCUITS
V.F. FROM V.F. PAD
V.F. TO TRANSMITTER
LOW PASS FILTER
IN
OUT
GRD
C
E
A
R
H
F
R-95 2500
Q-19 2N43
C-76 .5
R-98 51
CR23 IN765
C-43 .5µf
R-97 510
R-94 8200
R-96 3300
Q-18 2N43
CR27 IN765
R-90 6200
Q-17 2N43
C-42 .5µf
R-93 2K
R-89 5100
R-91 1K
R-67 1K
C-41 .5µf
Q-16 2N43
R-88 510
R-87 20K
R-85 22K
R-86 620
T.P. AMPL INPUT
V.F. TO V.F. RECEIVE PAD
J
P
"Z" WIRING STRAP
"X" WIRING STRAP
"Y" WIRING STRAP
CR26 IN765
BAND PASS FILTER 3.5KC
V.F. FROM RECEIVER
R-65 10K
R-64 5100
C-38 .068 ±3%
R-66 10K
Q-15 2N43
CR25 IN765
R-63 620
ST-9
R-61 620
R-60 1K
R-62 30K
C-39 .01
T.P. OSC OUTPUT
R-84 12K
R-69 5100
R-68 15K
K
TONE SIGNAL TO TRANSMITTER

Fig. 5. Signaling card

the hybrid are 0 dbm and −4 dbm respectively.

SIGNALING

With the equipment wired for "tone off during seizure" (*X* wiring option) the tone oscillator continually feeds a 3.5-kc tone during idle time into the modulator circuit of the transmitter card. This tone modulates the carrier at a 16 db lower level than the voice frequency.

Transmission of the tone modulated signal is the same as previously described.

The 3.5-kc output from the receiver at approximately −12 dbm level is fed into the signaling card through a 3.5-kc band-pass filter (see Fig. 4 for band-pass characteristics) to be amplified, clipped, and converted to d-c to energize the coil of the signaling relay.

As long as the *M* lead at the near end is open (or grounded) the 3.5-kc tone will be transmitted and the signaling relay will be energized at the far end and no ground will appear on the *E* lead bus.

When −48 volts is applied on the *M* lead input the oscillator is biased to cut off. Subsequently, the signaling relay at the far end drops out placing a ground on the *E* lead bus seizing the trunk equipment at that end.

Other options for signaling are as follows:

1. −24 volts on the *M* lead (use *Z* wiring Fig. 5).
2. Tone on during seizure (use *Y* wiring Fig. 5).

Receiver and transmitter cards have been discussed in other literature,[1-3] however because of the novel characteristics of the signaling circuit a brief description of its operation follows. See the schematic shown in Fig. 5.

It will be noted that the card has five distinct sections which are:

1. Oscillator.
2. Tone amplifier.
3. 3.5-kc tone band-pass filter.
4. 3-kc low-pass filter for the v-f transmit signal.
5. 3-kv low-pass filter for the v-f receive signal.

The oscillator is of the L-C type employing a single transistor. The oscillator is shown in the upper left area of Fig. 5. The tone band-pass filter is at the lower left center with the v-f receive 3-kc low-pass filter located to the right of it. The v-f transmit 3-kc low-pass filter is shown at the upper right.

The filters are those shown in Fig. 3 and are located on the signaling card to improve system flexibility, for convenience

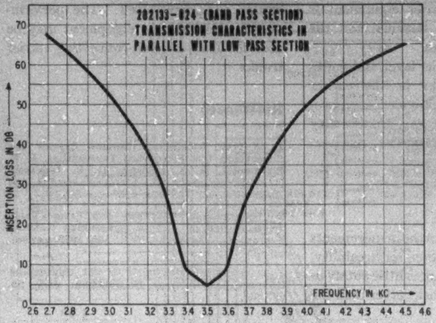


Fig. 6. Characteristics of 3.5-kc band-pass filter

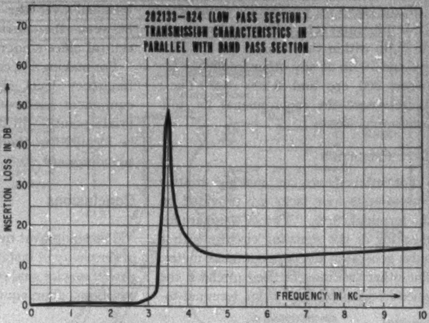


Fig. 7. Characteristics of 3-kc low-pass filter

of manufacture, and to reduce servicing problems.

For tone-on during idle, the tone oscillator transistor *Q*-15 is supplied with −48 volts through *R*64, *R*65, and transformer *ST*9. It will oscillate continuously until biased to cut off by the application of −24 to −48 volts on the *M* lead through *R*-67, *R*-66, and diode *CR*-25.

For tone-on during seizure, the oscillator receives all of its power fom the *M* lead and therefore will oscillate only during the time that voltage is supplied by it.

The 3,500-cps tone generated by the oscillator is fed to the transmitter at a level to produce sidebands 29 db below carrier level.

The signaling tone amplifier obtains its signal from the receiver through the 3.5-kc band-pass filter.

*Q*16 is a conventional common emitter amplifier. *Q*17 is an emitter follower which serves to match the output impedance of *Q*16 to the low input impedance of *Q*18. *Q*18 is a clipper to limit the magnitude of the signal before conversion to d-c in the last stage.

In order to understand the a-c-to-d-c conversion in the last stage, first consider it as an a-c amplifier with a controllable amount of feedback. As the feedback is increased the voltge available for conversion to d-c is reduced. The transistor may now be considered as a triode detector or a d-c amplifier. In either consideration the d-c output will be less as the a-c feedback is increased and will result in a slower operate time of the signaling relay. Rheostat *R*95 can therefore be used as a per-cent make control for dial pulses. It is possible to adjust the per-cent make of the signal receive relay to be equal to, greater than, or less than the per-cent make in the *M* lead at the transmitting terminal. The range of pulse conversion is ±4%.

In order to obtain reliable signaling and prevent talkdown, special characteristics were designed into the 3-kc low-pass filters and the 3.5-kc band-pass filter. These are as follows:

To insure perfect response to the 3.5-kc tone, the passband of the band-pass filter is 200 cycles wide. Outside the band the attenuation increases rapidly, particularly on the low-frequency side where the chances of talkdown are greater. Fig. 6 shows the attenuation characteristics of this filter.

The 3-kc low-pass filters also are of special design. They serve two purposes. The first requirement is to prevent 3.5-kc components of speech signals from entering the transmitter or signaling amplifier where they may cause eratic operation. The second requirement is to prevent the 3.5-kc signaling tone from getting out into the v-f line as an annoying whistle. Fig. 7 shows the attenuation of the 3-kc low-pass filter. Note the high attenuation characteristics of the filter at 3.5 kc. The use of these filters coupled with other special features of the system provide accurate signaling without pulse distortion in systems working over lines having losses exceeding 35 db at carrier frequencies (see Fig. 4). The filters in part make it possible to obtain F1A weighted noise levels lower than 15 dba (decibels adjusted).

The v-f band width of the system is 250 cps to 3,000 cps drop-to-drop with less than +1−3 db variation from the thousand cycle level. See Fig. 8 for drop-to-drop response curves of an actual installation at Merchants & Farmers Telephone Company of Montpelier, Va.

The design characteristics of the carrier frequency band-pass filters warrant some comment. Their passband approaches 3.5 kc each side of carrier in order to accommodate the sidebands of the signaling tone. They have an extremely flat passband, having a ripple of less than ±0.25 db.

The out-of-band attenuation characteristics are such that the channels may be spaced 12 kc apart at frequencies of 27, 39, ... 123, 135 kc without excessive crosstalk between channels.

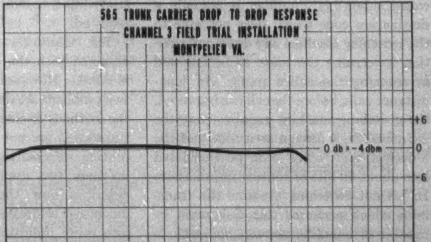


Fig. 8. Typical frequency response

Data taken on production equipments gave minimum return loss measurements of 22 db within the band of 500 to 2,500 cps and 19 db at any other frequency between 100 cps and 3,200 cps. These figures meet the requirements for toll-connecting trunks.

Test Facilities

For convenience in making routine system tests four jacks are provided on the terminal cards. Two of these are for checking of the signaling circuits with a pulse generator and per-cent make meter. The other two may be used to check the voice-frequency signal from terminal-to-terminal over the carrier equipment or between the carrier terminal and the central office equipment.

For making d-c, carrier level, and voice-frequency level checks in the carrier equipment, jacks are provided at appropriate points. In this way, trouble may be localized in the transmitter, receiver, signaling card, etc.

Conclusions

The equipment meets the requirements of simplicity of installation, serviceability, ease of service, and requires little outlay in additional plant equipment other than the carrier equipment itself. The performance is such that the equipment may be used for EAS and toll-connecting carrier trunks.

References

1. A New 5-Channel Carrier System for Subscriber Use Employing Transistors, B. R. Stachiewicz, R. L. Layburn, D. R. Jamieson. *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 76, Sept. 1957, pp. 385–90.

2. Engineering Features and Field Trial Performance of a New Subscriber Carrier System, R. L. Layburn. *Ibid.* (Jan. 1958 section), pp. 681–87.

3. Mechanical Features of 561 Subscriber Carrier, J. E. Ross, O. H. Goldman. *Ibid.*, pp. 677–81.

A Wired-Memory Translator with Shared Access

F. P. PACE
NONMEMBER AIEE

B. OSTENDORF, JR.
MEMBER AIEE

THIS PAPER describes the requirements, the circuits, and the design philosophy of a translator[1] developed in connection with experimental work on telephone and data-switching systems. The requirements called for a flexible translator of large capacity, with a nonvolatile memory, and for operation at speeds compatible with the present-day automatic telephone and telegraph switching systems. Moreover, the common translator must be made accessible to a large number of circuits which require translation service.

An important use of this translator is to convert destination-type codes or addresses into routing information. Two examples of this usage follow:

Example 1. In the telephone system a called subscriber's number is actually a destination code which does not give to the switching machine any direct information concerning the route to be followed in reaching the called party. In long-distance calls, before the call can be routed, the switching machine requires that the 3-decimal-digit foreign area code and the 3-decimal-digit office code be translated into about 60 bits of routing information. In present switching systems the translator which performs this task must be capable of handling about 1,200 such 6-digit codes. Also, the translator must be flexible in so far as code changes are concerned since the routing schemes are subject to frequent changes as the telephone traffic load changes. At present, this translation is handled by the card translator.[2]

Example 2. In certain data-switching systems messages are to be addressed by means of mnemonic codes which greatly facilitate the operator's task of memorizing many hundreds of station addresses. Also, in these systems group codes are used in such a way that a single message may be addressed to a large group of stations by means of a single group address. Again, there exists the need for translating these message addresses into route indicators which the switching machines can use for routing the messages to their proper destinations. In a switching system under study up to 2,000 such codes might be required.

The translator presented here was designed with the tasks described above in mind. However, some aspects of these two problems are different. For example, the methods of checking the results of the translation are not the same because the routing information is coded differently in these two applications. This paper describes how the translator converts mnemonic addresses and group addresses into route indicators for use by a data-switching system.

Consider the role of the translator functioning in a data-switching center, Fig. 1, in which the translator is designed to give service to a number of requesting machines. The translation of an address code into route indicators would proceed as follows: When a requesting machine has assembled the address to be translated it initiates a bid for translation service. This bid signal is detected by the translator, which operates an input access gate in the bidding requesting machine to allow the address which is to be translated to pass into the translator. Here it is converted into routing information. In the meantime, the translator also operates the output access gate in this requesting machine. In so doing, the results of the translation are transferred to the requesting machine. Then the translator proceeds to give service to the other bidding requesting machines in the order of their appearance on a scanner in the sequence circuit.

Translator Requirements

Translating Networks

The translating networks must be capable of accepting an address which consists of from two to eight alphabetical characters (five bits per character). Therefore, the input to the translator will be from 10 to 40 bits in increments of 5 bits.

Paper 60-126, recommended by the AIEE Data Communication Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., January 31–February 5, 1960. Manuscript submitted November 4, 1959; made available for printing December 4, 1959.

F. P. Pace is with the Airborne Instruments Laboratories, Melville, N. Y., and B. Ostendorf, Jr., is a consultant, Stamford, Conn.; both were formerly with the Bell Telephone Laboratories, Inc., New York, N. Y.

The authors wish to acknowledge the contributions of J. R. Davey, G. A. Locke, W. O. Fleckenstein, E. T. Ball, and E. R. Robinson to the systems and equipment aspects of the translator.

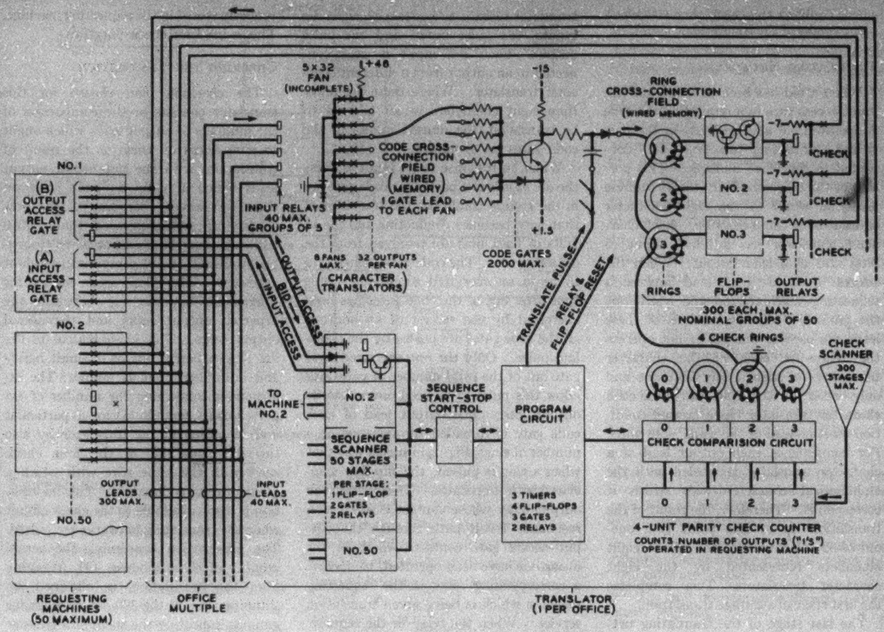

5×32
FAN
(INCOMPLETE)
+48
-15
RING
CROSS-CONNECTION
FIELD
(WIRED MEMORY)
NO. 1
(B)
OUTPUT
ACCESS
RELAY
GATE
(A)
INPUT
ACCESS
RELAY
GATE
CODE CROSS-CONNECTION FIELD (WIRED MEMORY)
1 GATE LEAD TO EACH FAN
INPUT RELAYS
40 MAX.
GROUPS OF 5
8 FANS MAX.
32 OUTPUTS PER FAN
(CHARACTER TRANSLATORS)
+1.5
CODE GATES
2000 MAX.
-7
CHECK
NO. 2
NO. 3
RINGS
FLIP-FLOPS
OUTPUT RELAYS
300 EACH, MAX.
NOMINAL GROUPS OF 50
4 CHECK RINGS
CHECK SCANNER
300 STAGES MAX.
OUTPUT ACCESS
BID
INPUT ACCESS
TRANSLATE PULSE
RELAY & FLIP-FLOP RESET
NO. 2
TO MACHINE NO. 2
SEQUENCE START-STOP CONTROL
PROGRAM CIRCUIT
SEQUENCE SCANNER
50 STAGES MAX.
PER STAGE:
1 FLIP-FLOP
2 GATES
2 RELAYS
3 TIMERS
4 FLIP-FLOPS
3 GATES
2 RELAYS
0 1 2 3
CHECK COMPARISION CIRCUIT
4-UNIT PARITY CHECK COUNTER
COUNTS NUMBER OF OUTPUTS ("1'S") OPERATED IN REQUESTING MACHINE
OUTPUT LEADS 300 MAX.
INPUT LEADS 40 MAX.
NO. 50
REQUESTING MACHINES (50 MAXIMUM)
OFFICE MULTIPLE
TRANSLATOR (1 PER OFFICE)

Fig. 1. Common translator system, functional diagram

The maximum number of addresses which the translator must be capable of translating was set at 2,000. It was also required that there be no restrictions placed on the way in which these addresses are chosen.

The output of the translating networks is 300 bits of routing information on 300 leads. Because of power supply limitations the maximum number of ONES on the output leads (operated output relays) was limited to 100.

The 40 input bits give about 10^{12} possible input codes. Of these almost any 2,000 codes may be chosen. These codes may be translated into any of 4×10^{82} possible output combinations. These large numbers of possible input and output combinations make it apparent that any co-ordinate type of network matrix which even approaches full implementation would require an astronomical number of crosspoints. This makes it necessary to design a simple network scheme which requires almost no equipment for codes which are merely possible but not actually used. Equipment is then added to the network only on a per code used basis (job wiring).

Common Usage

The requirements call for one translator to give translation service to as many as 50 requesting machines in a data-switching center. Because of this common usage a lockout function must be provided to permit the requesting machines to have orderly access to the translator. An electronic scanner was chosen to perform this function. Also, some input access equipment is required to control the flow of addresses to be translated from the requesting machine to the translator. Likewise, output access equipment is required to control the flow of routing information from the translator to the requesting machine. For compatibility with present-day plant the translator holding time requirement was set at 35 milliseconds per translation.

Auxiliary Functions

Besides the essential features described above, the translator must provide check circuitry to evaluate the results of each translation. Also, in case of circuit failure, spare equipment must be provided as replacement for the circuit which has failed. Some means are needed for monitoring and testing the equipment in service and testing out-of-service or spare circuits.

Translator Circuits and Operations

The translator consists of two essential elements as follows.

The translating networks, whose functions are:

1. To receive the address to be translated.
2. To furnish the routing information.

The sequence and program circuits, whose functions are:

1. To detect the bids for translation service.
2. To operate the input access gate of the bidding requesting machine to allow the address which is to be translated to pass into the translating networks.
3. To pulse the translating networks.
4. To operate the output access gate of the bidding requesting machine to allow the resulting routing information to pass from the translating networks to the requesting machine.

5. To direct the check circuit to check the results of the translation.

TRANSLATING NETWORKS

When a bid has been detected and the input access gate in a requesting machine is operated, 40 address leads in the particular requesting machine are connected to the office multiple. This portion of the office multiple (the 40 address leads) is wired to the eight character translators in the translator. Each character translator has a 5-lead input (5 bits) and a 32-lead output. When the address leads are connected between a requesting machine and the translator, the ground or open conditions of these leads are used to operate or not operate the relays in each of the character translators in such a way that one and only one of the 32 output leads from each character translator has a ground condition on it instead of the usual +48 volts. For convenience each output lead of a character translator is labeled with the alphabetical character with which it corresponds. Therefore, the result of the translation up to this point is a one-out-of-32 selection for each of the eight alphabets represented by the eight character translators. This completes the first stage of a 2-stage translation.

The last stage of the translating networks consists of up to 2,000 code gates, one code gate for each address which is used in the system. Each code gate is a transistor-resistor AND gate with eight d-c inputs, one pulse input, and one pulse output. Each d-c input is cross-connected to an output from a different character translator. Where there are fewer than eight characters in an address to be translated, the unused inputs to the code gates are grounded.

When, in response to an input address the d-c inputs to a particular code gate are in the ground condition, the transistor of that gate becomes conducting and the 15 volts of back bias are removed from the output diode. The code gate is then said to be in an energized condition. Some time after one of the code gates has been energized by the receipt of an address, all the code gates are pulsed by the translate pulse. Only the one energized code gate out of the total number of gates will allow this pulse to proceed onward to its output lead. The output lead of each code gate is threaded through one or a number of rings, Fig. 2, in such a way that when a ring is pulsed, the flip-flop associated with it operates. This operates an output relay whose contact closes a low-resistance 7-volt path through the output access gate contacts which in the meantime have been operated, to operate a corresponding relay in the requesting machine which is being given translation service. When the relay in the requesting machine operates, a ground is returned to the output relay contact as an indication that the information has been properly stored in the requesting machine. This is used for check purposes.

CHECKING THE TRANSLATION

The checking plan chosen for this translator consists of the comparison of the numbers of output leads which ought to have ONES on them as the result of translation, with the number of output leads which actually do have ONES on them. However, because of the low density of addresses used (only 2,000 addresses may be used out of 10^{12} possible addresses) a full comparison was not necessary. The plan used checks for congruence modulo (4) between the expected output ONES and the actual output ONES. This is equivalent to the use of two parity bits or a 4-unit parity and is implemented as follows: The information concerning the number of expected ONES modulo (4) for a particular code is stored in the translator by also threading one out of the four check rings with the same wire with which a group of rings is threaded. This information is made available to the check circuit when the translating networks are pulsed. The information concerning the actual number of ONES modulo (4) resulting from the translation is obtained by scanning (Fig. 3) the 300 output leads for grounds indicating the successful transfer of ONES to the requesting machines, and counting these grounds on a recycling 4-stage check counter. The check consists in matching the state of the four check ring flip-flops with the final state of the check counter. A match indicates a check.

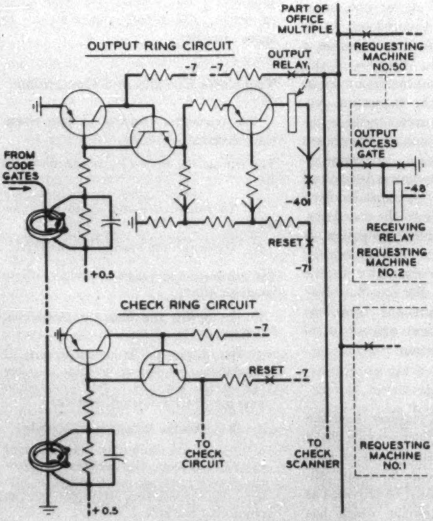


Fig. 2 (left). Ring operated flip-flops

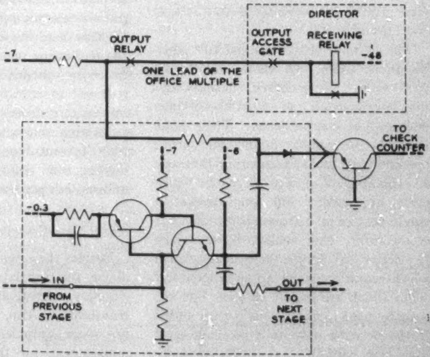


Fig. 3 (below). Check scanner stage

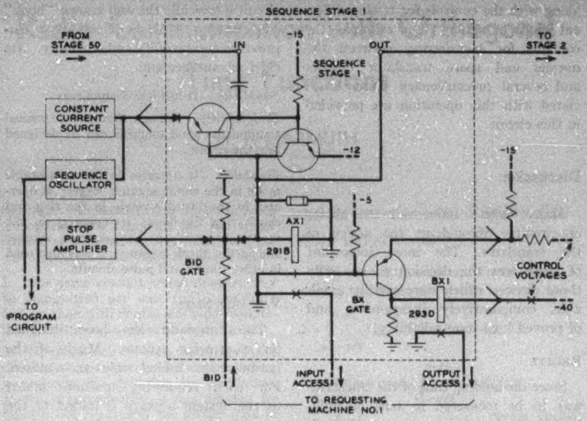


Fig. 4. Sequence scanner stage

Sequence Scanner

The purpose of the sequence scanner, Fig. 4, is to provide a means whereby the requesting machine may have orderly access to the translator with no preference given any of them. This task is accomplished by means of a 50-stage ring counter. Each stage is a series negative-resistance flip-flop. All stages have one terminal connected in common to a constant current supply which is sufficient to keep just one stage in the ON condition at a time. This ON condition has been designed to step along at a rate of 7,200 steps per second in response to pulses from a blocking oscillator (sequence oscillator) so long as the translator is not busy translating. When a stage is in the ON condition and bid gate of that stage is energized to detect the presence of a bid voltage (−48 volts) on the bid lead of the requesting machine associated with that particular stage of the sequence scanner. If a bid voltage is detected a STOP signal is sent to the program circuit. In turn, the program circuit stops the sequence oscillator and the ON condition remains at the stage where the bid signal was detected.

When the ON condition remains at a stage for more than 1 millisecond the mercury relay *AX* which is the load of that stage will operate. The *AX* make-contact operates the input access gate in the requesting machine, thus allowing the address which is to be translated to come into the translating networks. Also, the break contact of the *AX* relay is used to prime the *BX* gate so that, some time later, under control of the program circuit the *BX* relay of this stage will be operated. Operation of the *BX* relay operates the output access gate in the requesting machine and allows the routing information, which is the result of the translation, to proceed into the requesting machine.

Program Circuit

The various phases of a translation are directed by a wired-in program which consists of some flip-flops, timers, pulsers, and logic gates. A sequence chart shows when the various phases of translation occur with respect to one another; see chart of Fig. 5.

When the translator is not translating, the sequence scanner is scanning the 50-bid leads. As soon as a bid is detected an input access gate is operated; this operates the character translators, which in turn energize one of the code gates completing the encoding phase of the translation.

After sufficient time has elapsed to allow circuit transients to become negligible, all the code gates are pulsed. The output relays which have been assigned to this one energized code gate by means of the ring threadings are thus operated. In the meantime the output access gate of the requesting machine being given service has been closed, allowing the operating output relays to operate their associated relays in the requesting machine. This completes the translation phase of the over-all operation.

A time interval is again allowed for the circuit transients to disappear. At the end of this interval several operations occur simultaneously, as follows:

1. The translation results are checked and, after the check is completed, the output access gate of the requesting machine is released and the ring flip-flops, the check circuit, and the output relays are reset.
2. The input access gate of the requesting machine, the character translators and the code gates are released.
3. The sequence scanner is allowed to commence scanning.

If the next requesting machine is bidding, this will be immediately detected and the input access gate of this requesting machine will be operated. Thus the character translators and the code gates will be operated according to the new address to be translated, even while the output information is still being processed for the previous translation.

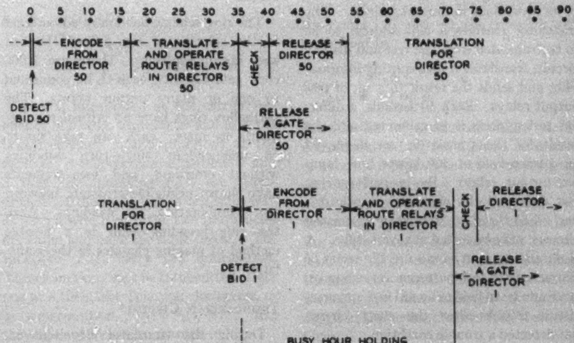


Fig. 5. Timing chart

At the time that the output portions of the translating networks have been reset, the input portions of the translating networks will have been energized so that the translator is ready to enter the translation phase of its operation for the second time.

Although a particular requesting machine is attached to the translator for about 53 milliseconds, by adopting the overlapping program described above the translator is able to complete a translation every 37 milliseconds. This gives to the translator an effective busy-hour holding time of 37 milliseconds.

DUPLICATED CIRCUITRY

Because this system is designed for real time operation those portions of the translator whose failure would result in total or nearly total communications breakdown were duplicated for quick substitution during emergencies. This includes the character translators, the program circuit, and the sequence circuit. These circuits can be transferred into service by operating a bank of transfer relays, the use of which also allows for routine exercising of the spare circuits.

TEST AND DISPLAY CIRCUIT

The test and display circuit contains the circuitry to perform the auxiliary functions described below.

By simulating portions of a requesting machine the test and display circuit is able to monitor the operation of the translator while it is in service. This is done as follows: Every 10 seconds, a test bid signal is initiated by the circuit. When this bid is detected by the sequence scanner, the translator proper proceeds in its operations in normal fashion. The translator receives a test address to be to be translated from the test and display circuit, translates it into routing information, and sends the result to a set of test output relays. Each 10 seconds, a different test address is so generated and the results of the translation are displayed on a lamp field of 300 lamps (one lamp per output relay). By properly wiring test codes all the character translators, and all of the rings with their associated circuits, may be tested automatically. A maintenance man would use the series of characteristic test patterns appearing on the route lamp field as an aid in diagnosing circuit trouble after the check circuit has detected a trouble condition.

Alarm features for indicating various types of circuit difficulties have been included in the test and display circuit along with the controls for testing different kinds of action. For example, the controls for transferring between the normal and spare translator circuits and several precautionary lamps associated with this operation are provided in this circuit.

Discussion

Many factors influenced the choice of circuitry throughout the design of the translator. The most important of these was the decision to use only those devices which were readily available, comparatively inexpensive, and of proved long-term reliability.

RELAYS

Since the holding time of the translator was to be measured in tens of milliseconds and not in microseconds, advantage could be taken of the more useful properties of relays. For example, the relay fan in the character translators originally consisted of about 130 diodes; these were replaced by eight relays and eight diodes. The relay fan is more economical and also has greater current-carrying capacity.

Diode gates were also considered for the input access gates and the output access gates which control the information flow to and from the translator. These gates would require about 35,000 diodes spread out over a 350-lead multiple which goes to 50 requesting machines. In view of the equipment costs and the vulnerability to diode damage on this complicated multiple, it was decided to use relays in this application also.

RINGS[3]

The ring scheme has many advantages in the solution of the difficult fan-out and fan-in problems: There are 2,000 code gate outputs, each to be connected to one or many output points. The permalloy rings have no critical requirements to meet, each can take up to 50 independent single-turn windings without crowding, and measurements have shown negligible crosstalk between any combination of windings. Also, the ring-threading plan is a flexible method of making changes in the translator.

TRUNCATED-N CIRCUIT[4]

During the translator development, a particularly interesting switching circuit was designed with an input characteristic which resembles a truncated N. This circuit is basically the well-known "hook" configuration with modifications to improve its designability and flexibility. Its chief advantages are:

Simplicity. It has few components.

Designability. Because there are no critical components good margins can be designed into the circuit.

Flexibility. It operates in the monostable mode in the check scanner stages; it operates in the bistable mode in the ring and check ring flip-flops; it operates in the multistable mode in the sequence scanner and in the check counter; it was also used in other timing and pulse circuits.

VARIABLE SIZE

The translator has been designed for a growing system. Much of the hardware is added only as required. For each requesting machine added to the system a stage is added to the sequence scanner. For each new address which the system has to handle a code gate is added to the translator. For each new bit required at the output of the translator a ring circuit and a check scanner must be added. When the message address format is expanded to include another character, a character translator is added. This growth aspect of the translator was needed to satisfy a variety of applications.

PRESENT STATUS

A laboratory "breadboard" model of the translator has been built in which all of the essential features of the translator have been tested and developed. A more complete model is being built for testing the circuits in their packaged environment and for evaluating the flexibility of the equipment arrangement for the ring circuits.

In conclusion, the translator is believed to be an economical balance of old art and new art techniques designed with sufficient flexibility to lend itself easily to the solution of various translation problems within the limitations of its requirements.

References

1. SOME BASIC CONCEPTS OF TRANSLATORS AND IDENTIFIERS USED IN TELEPHONE SWITCHING SYSTEMS, H. H. Schneckloth. *Bell System Technical Journal*, New York, N. Y., vol. 30, July, 1951, pp. 588–625.

2. THE CARD TRANSLATOR FOR NATIONWIDE DIALING, L. N. Hampton, J. B. Newsom. *Ibid.*, vol. 32, Sept. 1953, pp. 1037–98.

3. NO. 5. CROSSBAR AMA TRANSLATOR, T. L. Dimond. *Bell Laboratories Record*, New York, N. Y., vol. 29, Feb. 1951, p. 62.

4. A TRANSISTOR CIRCUIT WITH A TRUNCATED N CHARACTERISTIC, F. P. Pace. *Master's Thesis*, Polytechnic Institute of Brooklyn, Brooklyn, N. Y., June 1958.

Radio Communication with Orthogonal Time Functions

H. F. HARMUTH
ASSOCIATE MEMBER AIEE

Synopsis: Communication systems using orthogonal time functions for the transmission of information over radio links are characterized by the following properties:

1. The same amount of information may be transmitted per unit (pu) bandwidth and pu time with double-sideband modulation as with single-sideband modulation.
2. No phase lock between the carriers produced at the transmitter and at the receiver is necessary for detection by correlation. Hence, communication is less affected by rapid phase changes of the carrier than in systems requiring phase lock.
3. It is possible to operate under conditions of large Doppler shift without changing the frequency of the carrier produced at the receiver, provided the Doppler shift varies sufficiently slowly.

CONSIDER a set of time functions $F_s(t)$, $s=1, 2, \ldots m$ which are orthogonal in the interval $-\tau/2 \leqslant t \leqslant +\tau/2$.

$$E^{-1}\int_{-\tau/2}^{\tau/2} F_i(t)F_k(t)dt=\delta_{ik}=\begin{cases}1 \text{ for } i=k\\ 0 \text{ for } i\neq k\end{cases} \qquad (1)$$

Consider further the set of functions obtained by multiplying $F_s(t)$ by -1:

$$F_{-i}(t)=-F_i(t) \qquad (2)$$

Both sets of functions form an extended set of orthogonal functions:

$$E^{-1}\int_{-\tau/2}^{\tau/2} F_i(t)F_k(t)dt=\delta_{ik}' = \begin{cases}1 \text{ for } i=k\\ -1 \text{ for } i=-k\\ 0 \text{ for } i\neq k, -k\end{cases} \qquad (3)$$

where τ is finite if the functions $F_s(t)$ are time limited and infinite if they are not.

Let the carrier $\sqrt{2}\cos\omega_1(t+t_1)$ be amplitude modulated by $F_i(t)$,

$$f_i(t)=\sqrt{2}F_i(t)\cos\omega_1(t+t_1) \qquad (4)$$

and the carrier $\sqrt{2}\cos\omega_2(t+t_2)$ by $F_k(t)$:

$$f_k(t)=\sqrt{2}F_k(t)\cos\omega_2(t+t_2) \qquad (5)$$

The correlation of equations 4 and 5 during the interval τ yields:

$$I=\int_{-\tau/2}^{\tau/2} f_i(t)f_k(t)dt=I_1+I_2+I_3+I_4$$

$$I_1=\cos(\omega_1t_1-\omega_2t_2)\int_{-\tau/2}^{\tau/2}F_i(t)F_k(t)\times \cos(\omega_1-\omega_2)tdt$$

$$I_2=-\sin(\omega_1t_1-\omega_2t_2)\int_{-\tau/2}^{\tau/2}F_i(t)F_k(t)\times \sin(\omega_1-\omega_2)tdt$$

$$I_3=\cos(\omega_1t_1+\omega_2t_2)\int_{-\tau/2}^{\tau/2}F_i(t)F_k(t)\times \cos(\omega_1+\omega_2)tdt$$

$$I_4=-\sin(\omega_1t_1+\omega_2t_2)\int_{-\tau/2}^{\tau/2}F_i(t)F_k(t)\times \sin(\omega_1+\omega_2)tdt \qquad (6)$$

It may be shown by partial integration that I_3 and I_4 are negligibly small if $\omega_1+\omega_2$ are large compared with the highest frequency components of $F_i(t)$ and $F_k(t)$ that one wants to transmit. Integral I_2 vanishes if the two carrier frequencies ω_1 and ω_2 are equal.

If, in addition, the two carriers are in phase one obtains:

$$E^{-1}\int_{-\tau/2}^{\tau/2} f_i(t)f_k(t)dt = E^{-1}\int_{-\tau/2}^{\tau/2} F_i(t)F_k(t)dt=\delta_{ik}' \qquad (7)$$

If ω_1 is interpreted as the carrier frequency of a radio transmitter and ω_2 as the beat frequency locally produced at the receiver, it follows from equations 7 and 3 that the functions $F_s(t)$ are transmitted via the radio link in the same way as with a wire line, if the two frequencies are equal and have the same phase. The functions $F_s(t)$ carry the information. These functions may be transmitted either one at a time, like Teletype characters, or as superposition $F_1(t)+F_2(t)+\ldots$, like the digits of Teletype characters.

In the following the transmission of only one function, $F_i(t)$, will be discussed. Due to the orthogonality of the functions $F_i(t)$ the results remain unchanged if one substitutes a sum $\Sigma F_i(t)$ for $F_i(t)$.

Synchronous Communication

Consider two transmitters both using carriers with the frequency ω but with a phase shift of 90 degrees between the two carriers. The amplitude modulation with $F_i(t)$ or $F_l(t)$ yields:

$$f_i(t)=\sqrt{2}F_i(t)\cos\omega t$$

$$f_l'(t)=\sqrt{2}F_l(t)\sin\omega t \qquad (8)$$

Assume that the frequency ω can be produced exactly at the receiver but that there is a phase difference between the transmitter carrier and the receiver carrier. (The terms "transmitter carrier" and "receiver carrier" will be used to indicate where the carrier is produced.) The received signal $f_i(t)+f_l'(t)$ is thus correlated with the following functions:

$$f_k(t)=\sqrt{2}F_k(t)\cos(\omega t+\phi)$$

$$f_m'(t)=\sqrt{2}F_m(t)\sin(\omega t+\phi) \qquad (9)$$

Neglecting terms which are small for large values of ω, one obtains the following correlations between $f_i(t)+f_l'(t)$ and the functions 9 (the neglected terms are of the form of I_3 and I_4 in equation 6):

$$I_{ik}=E^{-1}\int_{-\tau/2}^{\tau/2}[f_i(t)+f_l'(t)]f_k(t)dt \doteq \delta_{ik}'\cos\phi-\delta_{lk}'\sin\phi$$

$$I_{lm}'=E^{-1}\int_{-\tau/2}^{\tau/2}[f_i(t)+f_l'(t)]f_m'(t)dt \doteq \delta_{lm}'\cos\phi+\delta_{im}'\sin\phi \qquad (10)$$

The first terms on the right in equation 10 represent the transmitted information. $\delta_{lk}'\sin\phi$ gives the crosstalk from "sine channel" to "cosine channel" and $\delta_{im}'\sin\phi$ the crosstalk from "cosine channel" to "sine channel." Since the δ' may assume the values 1, 0, and -1, the crosstalk will be defined by:

$$C=\frac{|\sin\phi|}{|\cos\phi|+|\sin\phi|} \qquad (11)$$

Introduction of numerical values yields a crosstalk of 10% for a phase difference of about ±6 degrees between transmitter and receiver carrier.

In order to be able to distinguish between sine channel and cosine channel at the receiver one may transmit a certain function $F_i(t)$, say $F_1(t)$, all the time through the cosine channel but never through the sine channel. A typical signal in the cosine channel has, then, the form:

$$f_i(t)=\sqrt{2}[F_i(t)+F_1(t)]\cos\omega t \qquad (12)$$

where $F_i(t)$ stands for any number of func-

Paper 60-142, recommended by the AIEE Communication Theory Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., January 31–February 5, 1960. Manuscript submitted June 11, 1959; made available for printing December 17, 1959.

H. F. HARMUTH is with the Stromberg-Carlson Company, Rochester, N. Y., a Division of General Dynamics Corporation.

The author expresses his gratitude to G. A. Franco and G. Lachs of the Stromberg-Carlson Company for several helpful suggestions.

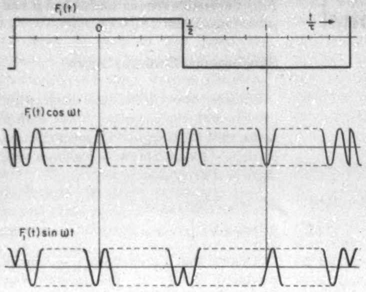


Fig. 1 (left). Modulation of cosine and sine carrier by square wave

tions of the set of equation 3. Since $F_1(t)$ is orthogonal to $F_i(t)$, $i\neq 1$, the integrals, equation 10, remain unchanged. Correlation of equation 12 with $\sqrt{2}\,F_1(t)$ sin $(\omega t+\phi)$ and $\sqrt{2}\,F_1(t)\cos(\omega t+\phi)$ yields:

$$I_{11}=2E^{-1}\int_{-\tau/2}^{\tau/2}[F_i(t)+F_1(t)]F_1(t)\times \cos\omega t\cos(\omega t+\phi)dt \doteq \cos\phi$$

$$I_{11}'=2E^{-1}\int_{-\tau/2}^{\tau/2}[F_i(t)+F_1(t)]F_1(t)\times \cos\omega t\sin(\omega t+\phi)dt \doteq \sin\phi \quad (13)$$

The output of the correlator producing integral I_{11}' may be used in a feedback loop to change the phase ϕ so that integral

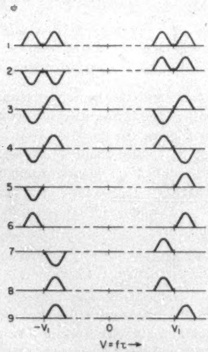


Fig. 3. Frequency spectra of symmetric and skew symmetric carrier and symmetric and skew symmetric modulation functions

1—$F_3(t)\cos\omega_1 t$
2—$F_3(t)\sin\omega_1 t$
3—$F_2(t)\cos\omega_1 t$
4—$F_2(t)\sin\omega_1 t$
5—$\frac{1}{2}[F_3(t)\sin\omega_1 t+F_2(t)\cos\omega_1 t]$
6—$\frac{1}{2}[F_3(t)\cos\omega_1 t-F_2(t)\sin\omega_1 t]$
7—$\frac{1}{2}[F_3(t)\sin\omega_1 t-F_2(t)\cos\omega_1 t]$
8—$\frac{1}{2}[F_3(t)\cos\omega_1 t+F_2(t)\sin\omega_1 t]$
9—$\frac{1}{2}[F_2(t)\,F_3(t)]\cos\omega_1 t$

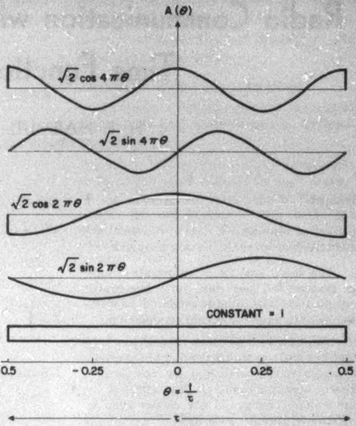


Fig. 2 (right). Set of orthogonal trigonometric time functions

I_{11}' vanishes. Information transmitted in the cosine channel will then be received in the cosine channel and the equivalent statement holds for the sine channel.

The integrals, 13, may be used to eliminate the crosstalk in equation 10 without feedback. Consider the correlation of $f_i(t)+f_i'(t)$ with the functions:

$$f_k'(t)=\sqrt{2}F_k(t)\sin(\omega t+\phi) \quad (14)$$

and

$$f_m(t)=\sqrt{2}F_m(t)\cos(\omega t+\phi)$$

$$I_{ik}'=E^{-1}\int_{-\tau/2}^{\tau/2}[f_i(t)+f_i'(t)]f_k'(t)dt = \delta_{ik}'\sin\phi+\delta_{lk}'\cos\phi$$

$$I_{lm}=E^{-1}\int_{-\tau/2}^{\tau/2}[f_i(t)+f_i'(t)]f_m(t)dt = -\delta_{lm}'\sin\phi+\delta_{im}'\cos\phi \quad (15)$$

A combination of equations 10, 13, and 15 yields δ_{ik}' and δ_{lm}' free of crosstalk:

$$I_{ik}I_{11}+I_{ik}'I_{11}'=\delta_{ik}'$$

$$I_{lm}'I_{11}-I_{lm}I_{11}'=\delta_{lm}' \quad (16)$$

Elimination of crosstalk by means of the phase information of equations 13 provides the same result as synchronization of the phase of the receiver carrier which would also be based on equations 13. Whether one prefers phase synchronization or elimination of the crosstalk with the help of equations 16 depends on how conveniently either method can be implemented under given conditions. There is, however, one important difference. The phase information obtained by the integration 13 during the time interval τ is used in equation 16 to eliminate crosstalk from the signals which were received during the same interval τ. If equation 13 is used to synchronize the receiver carrier, the phase information obtained during the interval τ will be used at a later time. This time lag makes phase synchronization undesirable if the phase of the transmitter carrier changes rapidly. Use of equations 13 for synchronization or crosstalk reduction by means of equations 16 is based on the assumption that phase shifts vary linearly with the frequency of the spectral components of $F_1(t)$ and $F_i(t)$.

It has been tacitly assumed so far that the period τ of the functions $F_i(t)$ is a multiple of the period of the carrier frequency. Modulation of a cosine or a sine carrier would then produce functions like $F_1(t)\cos\omega t$ and $F_1(t)\sin\omega t$ as shown by Fig. 1. To remove this undesirable requirement of synchronization between $1/\tau$ and ω, substitute $\cos\omega(t+t_1)$ and $\sin\omega(t+t_1)$ for $\cos\omega t$ and $\sin\omega t$. It follows from equations 6 for $\omega_1=\omega_2=\omega$ that t_1 and $t_2=t_1+\phi/\omega$ do not occur in the integrals. Furthermore, the functions $\cos\omega(t+t_1)$ and $\sin\omega(t+t_1)$ are only a different notation for $\cos(\omega t+\phi)$ and $\sin(\omega t+\phi)$. Only integrals I_3 and I_4 depend on the sum t_1+t_2 and these integrals vanish for large values of ω. Hence there is no need of making τ a multiple of $1/\omega$.

Single-Sideband Versus Double-Sideband Modulation

The following is a discussion of some specific sets of orthogonal functions. Fig. 2 shows the first five functions of a Fourier series. These functions can be produced by gating harmonic oscillators.

Vol.
Pt. 1
196

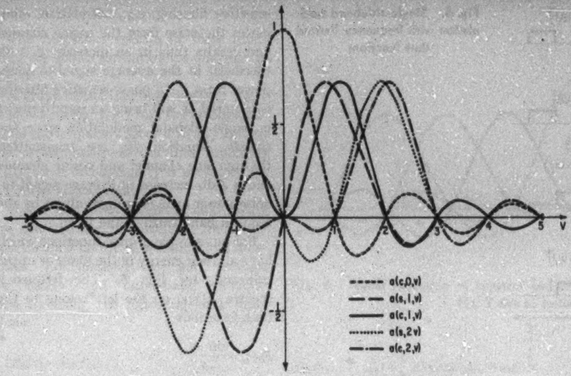


Fig. 4. Frequency spectra of trigonometric functions of Fig. 2

$$F_1(t) = -F_{-1}(t) = A$$

$$F_{2i}(t) = -F_{-2i}(t) = A\sqrt{2}\sin 2i\pi\theta$$

$$F_{2i+1}(t) = -F_{-2i-1}(t) = A\sqrt{2}\cos 2i\pi\theta$$

$$i = 1,2,3\ldots,\ \theta = t/\tau,\ -\frac{1}{2} \le \theta \le +\frac{1}{2} \qquad (17)$$

The frequency spectra of these functions have the following form:

$$a(c,0,\nu) = \frac{\sin \pi\nu}{\pi\nu}$$

$$a(s,i,\nu) = \frac{1}{2}\sqrt{2}\left[\frac{\sin \pi(\nu-i)}{\pi(\nu-i)} - \frac{\sin \pi(\nu+i)}{\pi(\nu+i)}\right]$$

$$a(c,i,\nu) = \frac{1}{2}\sqrt{2}\left[\frac{\sin \pi(\nu-i)}{\pi(\nu-i)} + \frac{\sin \pi(\nu+i)}{\pi(\nu+i)}\right] \qquad (18)$$

Graphs of these functions for $i=0$, 1, and 2 are presented in Fig. 2.

Modulation of the carrier $\sqrt{2}\cos\omega_1 t$ with the functions 17 produces the following frequency spectra ($\omega_1 = 2\pi f_1$, $\nu_1 = f_1\tau$):

$$a[\sqrt{2}F_1(t)\cos\omega_1 t,\nu]$$
$$= \frac{1}{2}\sqrt{2}[a(c,0,\nu-\nu_1)\underset{(-)}{+}a(c,0,\nu+\nu_1)]$$

$$a[\sqrt{2}F_{2i}(t)\cos\omega_1 t,\nu]$$
$$= (-)\frac{1}{2}\sqrt{2}[a(s,i,\nu-\nu_1)\underset{(-)}{+}a(s,i,\nu+\nu_1)]$$

$$a[\sqrt{2}F_{2i+1}(t)\cos\omega_1 t,\nu]$$
$$= \frac{1}{2}\sqrt{2}[a(c,i,\nu-\nu_1)\underset{(-)}{+}a(c,i,\nu+\nu_1)] \qquad (19)$$

The negative signs of the right members of equations 19 hold for the carrier $\sqrt{2}\sin\omega_1 t$.

To illustrate the effects of symmetry and skew symmetry of the modulating functions $F_{2i}(t)$, $F_{2i+1}(t)$ and the carrier the frequency spectra of $F_3(t)\cos\omega_1 t$, $F_3(t)\sin\omega_1 t$, $F_2(t)\cos\omega_1 t$, and $F_2(t)\sin\omega_1 t$ are plotted in Fig. 3.

It is seen from Fig. 4 that the frequency spectra of $F_{2i}(t)$ and $F_{2i+1}(t)$ are almost equal except that the 1's are skew symmetric and the others symmetric. Skew symmetry and symmetry becomes, for large values of i, the only difference between the frequency spectra. Sum or difference of the frequency spectra of $F_{2i}(t)$ and $F_{2i+1}(t)$ will have large values for $\nu>0$ or $\nu<0$ only. The orthogonal time functions belonging to these new frequency spectra are obtained by forming the sum and the difference of the functions $F_{2i}(t)$ and $F_{2i+1}(t)$.

$$F_i'(t) = \frac{1}{2}[F_{2i}(t) + F_{2i+1}(t)] = A\sin 2\pi\left(i\theta + \frac{1}{8}\right)$$

$$F_i''(t) = \frac{1}{2}[F_{2i}(t) - F_{2i+1}(t)]$$
$$= A\sin 2\pi\left(i\theta - \frac{1}{8}\right)$$

Frequency spectra $a[F_i'(t),\nu]$ of $F_i'(t)$ are shown in Fig. 5 for several values of i.

$$a[F_i'(t),\nu] = \frac{1}{2}\sqrt{2}\,\frac{\sin \pi(\nu-i)}{\pi(\nu-i)}$$

Modulation of the carrier $\sqrt{2}\cos\omega_1 t$ by $F_1'(t)$ yields the following spectra:

$$a[\sqrt{2}F_i'(t)\cos\omega_1 t,\nu]$$
$$= \frac{1}{2}\left[\frac{\sin \pi(\nu-\nu_1-i)}{\pi(\nu-\nu_1-i)} \underset{(-)}{+} \frac{\sin \pi(\nu+\nu_1-i)}{\pi(\nu+\nu_1-i)}\right]$$

The negative sign holds for the carrier $\sqrt{2}\sin\omega_1 t$.

Spectra of single-sideband modulation is not represented by $a[\sqrt{2}\ F_i'(t)\cos\omega_1 t,\ \nu]$ since the sign of i in the second term is the same as in the first term. The spectrum $a[\sqrt{2}\ F_i'(t)\cos\omega_1 t,\ \nu]$ belonging to the function $1/2[F_2(t)+F_3(t)]\cos\omega_1 t$ is shown in Fig. 3.

Single-sideband modulation is obtained by combining the function 17 with the carriers $\sqrt{2}\cos\omega_1 t$ and $\sqrt{2}\sin\omega_1 t$ in the following ways.

$$f_{2i}{}^{+}(t) = \frac{1}{2}\sqrt{2}[F_{2i+1}(t)\cos\omega_1 t - F_{2i}(t)\sin\omega_1 t]$$

$$f_{2i+1}{}^{+}(t) = \frac{1}{2}\sqrt{2}[F_{2i+1}(t)\sin\omega_1 t +$$
$$F_{2i}(t)\cos\omega_1 t]$$

$$f_{2i}{}^{-}(t) = \frac{1}{2}\sqrt{2}[F_{2i+1}(t)\cos\omega_1 t + F_{2i}(t)\sin\omega_1 t]$$

$$f_{2i+1}{}^{-}(t) = \frac{1}{2}[F_{2i+1}(t)\sin\omega_1 t -$$
$$F_{2i}(t)\cos\omega_1 t] \qquad (20)$$

Using the notation of equations 18 one may write the frequency spectra of the functions 20 in the following form:

$$a[f_{2i}{}^{+}(t),\nu] = a(c,i+\nu_1,\nu)$$

$$a[f_{2i+1}{}^{+}(t),\nu] = a(s,i+\nu_1,\nu)$$

$$a[f_{2i}{}^{-}(t),\nu] = a(c,i-\nu_1,\nu)$$

$$a[f_{2i-1}{}^{-}(t),\nu] = a(s,i-\nu_1,\nu) \qquad (21)$$

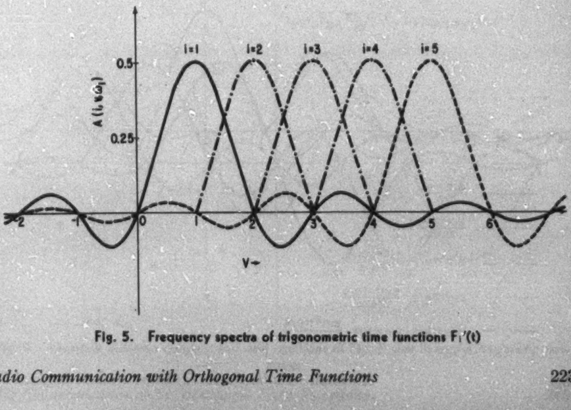


Fig. 5. Frequency spectra of trigonometric time functions $F_i'(t)$

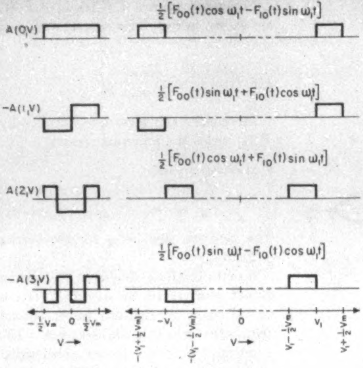

$\frac{1}{2}[F_{00}(t)\cos\omega_1 t - F_{10}(t)\sin\omega_1 t]$
A(0,V)
$\frac{1}{2}[F_{00}(t)\sin\omega_1 t + F_{10}(t)\cos\omega_1 t]$
−A(1,V)
$\frac{1}{2}[F_{00}(t)\cos\omega_1 t + F_{10}(t)\sin\omega_1 t]$
A(2,V)
$\frac{1}{2}[F_{00}(t)\sin\omega_1 t - F_{10}(t)\cos\omega_1 t]$
−A(3,V)
$-\frac{1}{2}V_m$
0
$\frac{1}{2}V_m$
V →
$-(V_1+\frac{1}{2}V_m)$
$-V_1$
$-(V_1-\frac{1}{2}V_m)$
0
V →
$V_1-\frac{1}{2}V_m$
V_1
$V_1+\frac{1}{2}V_m$

Fig. 6. Single-sideband modulation with frequency limited time functions

The functions 21 represent frequency spectra of single-sideband modulation with $f_{2i}^+(t)$ and $f_{2i+1}^+(t)$ having energy practically in the upper sideband only, and $f_{2i+1}^-(t)$ and $f_{2i}^-(t)$ in the lower sideband only. The spectra, equation 20, are shown in Fig. 3 for $i=1$. A comparison of equation 20 with equation 8 shows that single-sideband modulation is a practical implementation of the concept of transmission through sine channel and cosine channel. The functions $f_{2i}^+(t)$ and $f_{2i+1}^+(t)$ contain frequencies which occupy the same band but have a phase difference of 90 degrees. Hence, single-sideband modulation requires the same bandwidth for the same information rate as double-sideband modulation if the double-sideband modulation is performed in the way discussed.[1] It follows from equations 20 that single-sideband modulation may be performed by using two carriers with the same frequency but a phase difference of 90 degrees.[2,3] A double-sideband transmitter transmits all its energy either through the sine channel or the cosine channel of a certain frequency band; a single-sideband transmitter transmits its energy through the sine and cosine channel of half this frequency band; the power transmitted per bandwidth and per-phase channel is the same. As discussed in the Appendix, the components of white, normal noise may have the same frequency but a phase difference of 90 degrees and are then statistically independent.[4] If phase-insensitive filtering is used in a system which uses double-sideband modulation and transmits through the sine channel, the noise coming through the sine channel as well as through the cosine channel distorts the signal. Phase-sensitive filtering, e.g., correlation, eliminates the noise from the cosine channel and results thus in an increase of 3 db (decibels) in the average signal-to-noise-power ratio. No phase-sensitive filtering is required or will bring an improvement in single-sideband modulation since the signals, equation 20, are transmitted through sine channel and cosine channel and a 3-db increase in average signal-to-noise-power ratio is brought about by the reduced bandwidth of the filter.

For an example of time functions which have all their energy in the lower or upper sideband, see Fig. 6. The frequency spectra $A(n,\nu)$ on the left belong to the time functions

$$F_{00}(t)=\frac{\sin \pi\theta}{\pi\theta}$$

$$F_{10}(t)=\frac{\sin^2 \pi\theta/2}{\pi\theta/2}$$

$$F_{20}(t)=-\sin\frac{1}{2}\pi\theta\,\frac{\sin^2 \pi\theta/4}{\pi\theta/4}$$

$$F_{30}(t)=\cos\frac{1}{2}\pi\vartheta\,\frac{\sin^2 \pi\vartheta/4}{\pi\theta/4}$$

$$F_{40}(t)=-\cos\frac{1}{2}\pi\theta\,\sin\frac{1}{4}\pi\theta\,\frac{\sin^2 \pi\theta/8}{\pi\theta/8}$$

$$\vdots \qquad (22)$$

The following time functions with energy in the upper or lower sideband only may be derived from equation 22:

$$\frac{1}{2}\,[F_{00}(t)\sin\omega_1 t+F_{10}(t)\cos\omega_1 t]$$

$$\frac{1}{2}\,[F_{00}(t)\cos\omega_1 t-F_{10}(t)\sin\omega_1 t]$$

$$\frac{1}{2}\,[F_{20}(t)\sin\omega_1 t+F_{30}(t)\cos\omega_1 t]$$

$$\frac{1}{2}\,[F_{20}(t)\cos\omega_1 t-F_{30}(t)\sin\omega_1 t]$$

$$\vdots$$

$$\frac{1}{2}\,[F_{00}(t)\sin\omega_1 t-F_{10}(t)\cos\omega_1 t]$$

$$\frac{1}{2}\,[F_{00}(t)\cos\omega_1 t+F_{10}(t)\sin\omega_1 t] \qquad (23)$$

$$\vdots$$

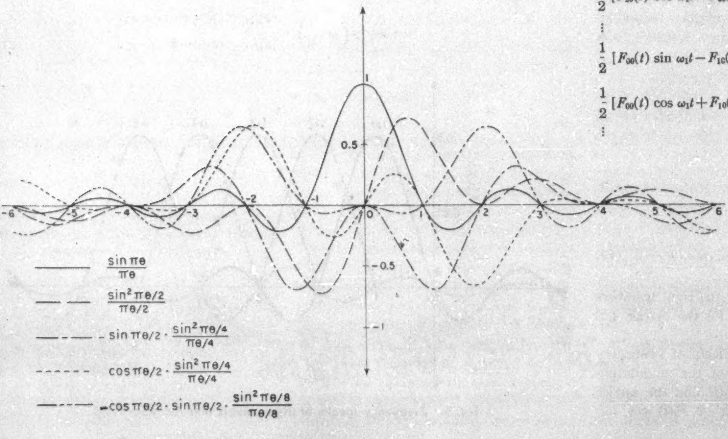


Fig. 7. Frequency limited time functions

The frequency spectra of these functions are shown on the right-hand side in Fig. 6.

A graph of the functions $F_{00}(t)$ to $F_{40}(t)$ is shown in Fig. 7. From the orthogonality of the frequency spectra $A(n,\nu)$ follows the orthogonality of the time functions $F_{n\nu}(t)$ since orthogonality is preserved by a Fourier transformation.

Upper and lower sideband modulation are special cases of arranging frequency spectra by a choice of suitable time functions. For instance, one may produce an "inner-" and "outer-band" modulation by using the following time functions:

$$\frac{1}{2}[F_{00}(t)-F_{20}(t)]$$

$$\frac{1}{2}[F_{10}(t)-F_{30}(t)]$$

$$\vdots$$

$$\frac{1}{2}[F_{00}(t)+F_{20}(t)]$$

$$\frac{1}{2}[F_{10}(t)+F_{30}(t)]$$

$$\vdots$$

The respective frequency spectra contain energy either in the inner band $-\nu_m/4 \leqslant \nu \leqslant +\nu_m/4$ or in the outer band $\nu_m/4 \leqslant |\nu| \leqslant \nu_m/2$ only, as may be inferred from Fig. 6.

Asynchronous Communication

If the frequencies ω_1 and ω_2 in equations 4 and 5 are not equal, the values of the integrals I_1 and I_2 in equations 6 depend on the particular set of orthogonal functions $F_*(t)$ used. Two sets of orthogonal functions will be investigated to outline the effects of frequency variations.

Consider first the set of trigonometric functions 17. Integral I_2 taken over the symmetric products $F_1(t)F_1(t)$, $F_1(t)F_{2i+1}(t)$, $F_{2i}(t)F_{2k}(t)$, and $F_{2i+1}(t)F_{2k+1}(t)$ equals zero; similarly, I_1 vanishes for the products $F_1(t)F_{2i}(t)$ and $F_{2i}(t)F_{2k+1}(t)$.

$$I_1=I_1(F_1,F_1)=\cos(\omega_1t_1+\omega_2t_2)\int_{-\tau/2}^{\tau/2}A^2\times \cos(\omega_1-\omega_2)tdt=A^2\tau\cos(\omega_1t_1-\omega_2t_2)\frac{\sin\pi\mu}{\pi\mu}$$

$$I_1(F_1,F_{2i+1})=-\frac{1}{2}\sqrt{2}A^2\tau\times \cos(\omega_1t_1-\omega_2t_2)\frac{\mu}{i}S(\mu+i)$$

$$I_1(F_{2i},F_{2k})=-\frac{1}{2}A^2\cos(\omega_1t_1-\omega_2t_2)\times \left[\frac{\mu}{i+k}S(\mu+i+k)+\frac{\mu}{i-k}S(\mu+i-k)\right]$$

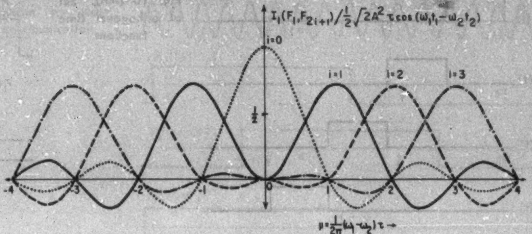


Fig. 8. Signal reduction of constant and crosstalk between constant and even functions of Fig. 2 due to carrier frequency variations

$$I_1(F_{2i+1},F_{2k+1})=\frac{1}{2}A^2\tau\cos(\omega_1t_1-\omega_2t_2)\times \left[\frac{\mu}{i+k}S(\mu+i+k)-\frac{\mu}{i-k}S(\mu+i-k)\right]$$

$$I_2(F_1,F_{2i})=\frac{1}{2}\sqrt{2}A^2\tau\cos(\omega_1t_1-\omega_2t_2)S(\mu+i)$$

$$I_2(F_{2i},F_{2k+1})=A^2\tau\sin(\omega_1t_1-\omega_2t_2)[S(\mu+i+k)+S(\mu+i-k)]$$

$$\mu=\frac{1}{2\pi}(\omega_1-\omega_2)\tau$$

$$S(\mu+i+k)=\frac{\sin\pi(\mu+i+k)}{\pi(\mu+i+k)}-\frac{\sin\pi(\mu-i-k)}{\pi(\mu-i-k)} \qquad (24)$$

Plots of the functions $I_1(F_1, F_{2i+1})$ are shown in Fig. 8. It may be seen that a frequency difference $\mu=\pm 1/2$ results in a crosstalk $I_1(F_1, F_3)$ from $F_3(t)$ to $F_1(t)$ which is almost as strong as the desired information $I_1(F_1, F_1)$.

For a Teletype system the period τ is the time required to transmit one character, $\tau=166.6$ ms (millisecond). The normalized frequency difference $\mu=\pm 1/2$ corresponds in this case to an actual frequency difference $f_1-f_2=\tau/\mu=\pm 3$ cps (cycles per second).

Time functions $F_i(t)$ which permit larger frequency differences may be obtained by transmitting, for instance, the function $F_1(t)+F_3(t)$ instead of $F_1(t)$. The permissible frequency difference μ is, in this case, increased to about $\mu\pm 3/2$, but the bandwidth required for a certain information rate is also increased.

Of more practical interest is the change of sign of $I_2(F_1, F_2)$ at $\mu=0$ in Fig. 9 which permits automatic tuning of the receiver carrier to the transmitter carrier.

The second set of orthogonal time functions to be considered is shown in Fig. 10. Superposition of these functions and the functions $F_{b,-i}(t)=-F_{b,i}(t)$, $i=1$, 2, 3, 4, 5, yields the 32 time functions produced by a balanced teletypewriter. The "worst" superpositions of these funtions, e.g., $F_{b1}(t)+F_{b2}(t)-F_{b3}(t)+F_{b4}(t)+F_{b5}(t)$ require about five times the bandwidth (measured by the first zero crossing of the frequency spectra) of any superposition of the five functions of Fig. 2. In return, these functions do not pro-

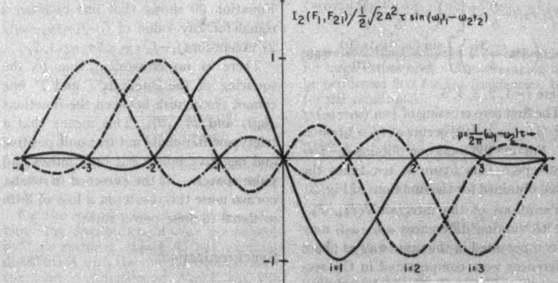


Fig. 9. Crosstalk between constant and odd functions of Fig. 2 due to carrier frequency variations

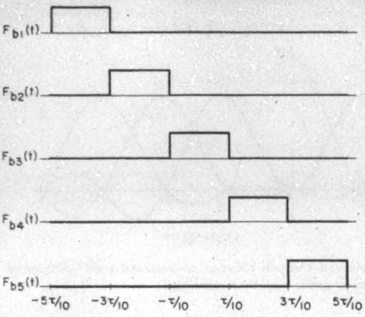


Fig. 10 (left). Set of orthogonal time functions

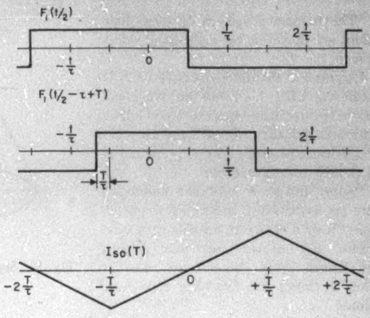


Fig. 11 (right). Synchronization for wire communication

duce any crosstalk due to differences of phase or frequency between transmitter and receiver carrier; peak power and average power are the same for any superposition of these functions whereas the superposition of the five functions of Fig. 2 produces peak powers which equal almost five times the average power. Integrals I_1 and I_2 of equations 6 may be readily computed. One obtains, for instance:

$$I_1=I_1(F_{b3},F_{b2})=\cos(\omega_1 t_1-\omega_2 t_2)\times \int_{-\tau/10}^{\tau/10}\cos(\omega_1-\omega_2)t\,dt=\frac{\tau}{5}\times \cos(\omega_1 t_1-\omega_2 t_2)\,\frac{\sin(\omega_1-\omega_2)\tau/10}{(w_1-w_2)\tau/10}$$

$$I_2=I_2(F_{b3},F_{b3})=-\sin(\omega_1 t_1-\omega_2 t_2)\times \int_{-\tau/10}^{\tau/10}\sin(\omega_1-\omega_2)t\,dt=0 \quad (25)$$

There are no crosstalk integrals.

$$I_1(F_{bi},F_{bk})=I_2(F_{bi},F_{bk})=0,\ i\neq k$$

Of interest are the sums

$$I(F_{bi},F_{bi})=-I(F_{bi},F_{b,-i})=I_1(F_{bi},F_{bi})+I_2(F_{bi},F_{bi})$$

$$I(F_{bi},F_{bi})=\frac{\tau}{5}\cos\left[\omega_1\left(t_1+\frac{i-3}{5}\tau\right)-\omega_2\left(t_2-\frac{i-3}{5}\tau\right)\right]\frac{\sin(\omega_1-\omega_2)\tau/10}{(\omega_1-\omega_2)\tau/10} \quad (26)$$

where $i=1, 2, 3, 4, 5$.

The first zero crossing of $[\sin(\omega_1-\omega_2)\tau/10]/[(\omega_1-\omega_2)\tau/10]$ occurs at $(f_1-f_2)\tau=\pm 5$. For $\tau=166.6$ ms one obtains $f_1-f_2=\pm 30$ cps. This figure is ten times the value obtained for the functions of Fig. 2.

Variations of the integrals $I(F_{bi}, F_{bi})$ due to varying differences $\omega_1 t_1-\omega_2 t_2$ may be compensated in the same way as phase differences were compensated in the section "Synchronous Communications." Let the received signal, equation 4, be correlated with:

$$f_k'(t)=\sqrt{2}F_k(t)\sin\omega_2(t+t_2) \quad (27)$$

One obtains:

$$I'=\int_{-\tau/2}^{\tau/2} f_i(t)f_k'(t)dt=I_1'+I_2'+I_3'+I_4'$$

$$I_1'=I_1'(F_i,F_k)=-\sin(\omega_1 t_1-\omega_2 t_2)\times \int_{-\tau/2}^{\tau/2}F_i(t)F_k(t)\cos(\omega_1-\omega_2)t\,dt$$

$$I_2'=I_2'(F_i,F_k)=-\cos(\omega_1 t_1-\omega_2 t_2)\times \int_{-\tau/2}^{\tau/2}F_i(t)F_k(t)\sin(\omega_1-\omega_2)t\,dt \quad (28)$$

I_3' and I_4' are of the form of I_3 and I_4 in equations 6 and vanish for large values of $\omega_1+\omega_2$.

The sums:

$$I'(F_{bi},F_{bi})=-I'(F_{bi},F_{b,-i})=I_1'(F_{bi},F_{bi})+I_2'(F_{bi},F_{bi}) \quad (29)$$

yield the values equations 26 except that cos in front of the bracket is replaced by −sin.

One obtains from equations 26 and 29:

$$I^2(F_{bi},F_{bi})+I'^2(F_{bi},F_{bi})=I^2(F_{b,i}F_{b,-i})+I'^2(F_{bi},F_{b,-i})=\left(\frac{\tau}{5}\right)^2\left(\frac{\sin(\omega_1-\omega_2)\tau/10}{(\omega_1-\omega_2)\tau/10}\right)^2 \quad (30)$$

Equation 30 shows that one receives a signal for any value of $f_1-f_2=(\omega_1-\omega_2)/2\pi$ except for $(f_1-f_2)\tau=\pm 5n$; $n=1, 2, \ldots$

There is no crosstalk. Due to the squaring of the integrals I and I' one cannot distinguish between the functions $F_{bi}(t)$ and $F_{b,-i}(t)$. This means that a teletypewriter could not transmit positive and negative pulses but only pulses and pulse spaces. In the presence of white, normal noise this results in a loss of 3 db in signal-to-noise-power ratio.

Synchronization

Orthogonal time functions may be constructed which permit a certain amount of time shift between $F_i(t)$ and $F_k(t)$ without invalidating relation 3. In general, however, one will try to avoid such a shift and synchronize $F_i(t)$ and $F_k(t)$. The principle of a synchronization method based on correlation will be explained with reference to Fig. 11.

The function $F_1(t/2)$, which has twice the duration of $F_1(t)$ in equation 17, and the function $F_{-1}(t/2)=-F_1(t/2)$ are transmitted alternatingly. At the receiver the same square wave delayed by τ is produced. Multiplication of the two square waves and integration from $-\tau/2$ to $+3\tau/2$ yields zero. If, however, the square wave of the receiver is advanced, the integration will produce a positive output; on the other hand, a negative output will be obtained if the receiver is retarded. I_{so} represents the output of the integrator for any time difference between the two square waves. The point $T/\tau=0$ is stable and the points $T/\tau=\pm 2$ are unstable if a positive value of I_{so} is used to retard the receiver and a negative one to advance it.

The information carrying functions $F_{2i}(t)$ and $F_{2i+1}(t)$ of equation 17 are orthogonol to $F_1(t/2)$, for $T=0$ they are also orthogonal to $F_1(t/2-\tau+T)$. A synchronization error produces crosstalk between the information carrying functions and $F_1(t/2-\tau+T)$. Hence, the integral I_{so} should be averaged over several periods τ to prevent the crosstalk from affecting the synchronization.

For the discussion of the application of the synchronization technique to radio communication refer to Fig. 12. The transmitter carrier $\cos\omega t$ is modulated by the square wave $F_0(t/2)$. At the receiver the two carriers $\cos(\omega t-\tau+T)$ and $\sin(\omega t-\tau+T)$ are modulated by the functions $F_0(t/2-\tau+T)$ and $F_0(t/2+T)$. One obtains the following correlations:

$$I_{s1}(T)=\int_{-(\tau/2+\tau)}^{3\tau/2-\tau}F_0(t/2)F_0(t/2-\tau+T)\times \cos\omega t\cos\omega(t-\tau+T)dt\doteq T\cos\omega T$$

$$I_{s2}(T)=\int_{-(3\tau/2+T)}^{\tau/2-T} F_0(t/2)F_0(t/2+T)\times$$
$$\cos \omega t \cos \omega(t-\tau+T)dt \doteq (\tau-T) \cos \omega T$$

$$I_{s3}(T)=\int_{-(\tau/2+T)}^{2\tau/2-T} F_0(t/2)F_0(t/2-\tau+T)\times$$
$$\cos \omega t \sin \omega(t-\tau+T)dt \doteq T \sin \omega T$$

$$I_{s4}(T)=\int_{-(3\tau/2+T)}^{\tau/2-T} F_0(t/2)F_0(t/2+T)\times$$
$$\cos \omega t \sin \omega(t-\tau+T)dt \doteq (\tau-T) \sin \omega T \quad (31)$$

$$I_{s1}(T)I_{s2}(T)+I_{s3}(T)I_{s4}(T) \doteq T(\tau-T);$$
$$0 \leqslant T \leqslant \tau \quad (32)$$

The sum 32 for any value of T is shown in Fig. 12. This function has only one half the period of the function $I_{s0}(T)$ in Fig. 11 but may otherwise be used in the same way for synchronization. For the phase synchronization of the receiver the function $I_{s3}(T)$ is better suited than the function 32 since it is subject to fewer mathematical operations.

Appendix

The following is a discussion of error probability due to white, normal noise.

Let the function $F_m(t)$ in the interval $-\tau/2 \le t \le \tau/2$ be represented by a superposition of the functions 17; that is, by a Fourier series.

$$F_m(t)=\sum_{k=-\infty}^{\infty} a_{mk}e^{2ik\pi t/\tau} \quad (33)$$

$$a_{mk}=\frac{1}{\tau}\int_{-\tau/2}^{\tau/2} F_m(t)e^{-2ik\pi t/\tau}dt$$

$$\int_{-\tau/2}^{\tau/2} F_m(t)F_m{}^*(t)dt=\tau\sum_{k=-\infty}^{\infty} a_{mk}a_{mk}{}^*=E \quad (34)$$

Consider white noise in the frequency domain. Let the frequency band be divided into intervals $\Delta f=f_l-f_{l-1}$. The probability that a sample of noise has a frequency component between f_l and f_{l-1} is independent of l. The amplitudes of the frequency components in any interval have a normal distribution. A sample of noise can be represented by a Fourier series:[4]

$$g(t)=\sum_{l=-\infty}^{\infty} A_l e^{2il\pi\Delta f t} \quad (35)$$

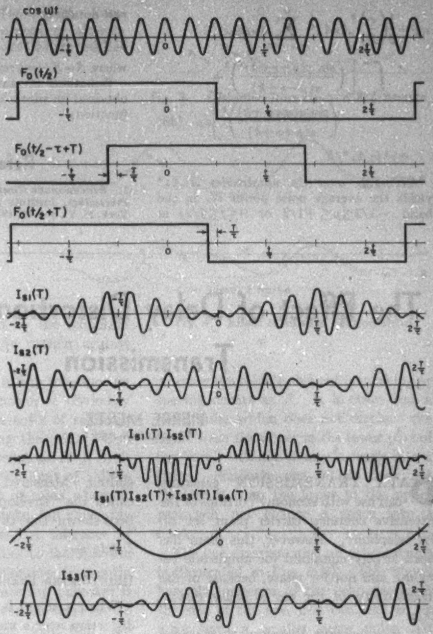
where the A_l have a normal distribution.

If a signal $\phi(t)$ is transmitted through a noisy channel it arrives at the receiver in the form $\phi(t)+g(t)$. The correlation of this signal with $\sqrt{2}F_m(t)\cos(\omega t+\phi)$ yields:

$$\sqrt{2}\int_{-\tau/2}^{\tau/2}[\phi(t)+g(t)]F_m(t)\cos(\omega t+\phi)dt$$

The contribution $N_m\tau$ of the noise to this integral is obtained with the help of equations 33 and 34.

Fig. 12. Synchronization for radio communication



$$N_m\tau=\sqrt{2}\int_{-\tau/2}^{\tau/2} g(t)F_m(t)\cos(\omega t+\phi)dt$$
$$=\frac{1}{2}\sqrt{2}\tau\sum_{k=-\infty}^{\infty}\sum_{d=-\infty}^{\infty} A_l a_{mk}\times$$
$$\{\cos\phi\, S(\mu-\nu+k)-i\sin\phi S'(\mu-\nu+k)\}$$

$$\mu=\tau l\Delta f, \nu=\tau f$$

$$S(\mu-\nu+k)=\frac{\sin \pi(\mu-\nu+k)}{\pi(\mu-\nu+k)}+\frac{\sin \pi(\mu+\nu+k)}{\pi(\mu+\nu+k)}$$

$$S'(\mu-\nu+k)=\frac{\sin \pi(\mu-\nu+k)}{\pi(\mu-\nu+k)}-\frac{\sin \pi(\mu+\nu+k)}{\pi(\mu+\nu+k)} \quad (36)$$

where μ becomes a continuous variable for $\Delta f \rightarrow 0$ and the sum over l becomes an integral.

For the computation of the error probability the distribution of $N_m\tau$ normalized by E is required. Since A_l has a normal distribution with the average equal to zero the joint distribution of $N_l\tau/E$ has the form:

$$W(N_l\tau/E)=(2\pi)^{-1/2}\sigma^{-1}\times \exp[-(N_l\tau/E)^2/2\sigma^2]$$

A component of $g(t)$ with amplitude A_λ and frequency $\mu=\lambda\Delta f\tau$ correlated with $\sqrt{2}F_m(t)\cos(\omega t+\phi)$ produces the term for $l=\lambda$ in equation 36. The square of this term divided by E is the square deviation $\sigma_{A\mu}{}^2$ for a particular amplitude and a particular frequency from the average $\overline{N_l\tau/E}=0$.

$$\sigma_{A\mu}{}^2=\frac{1}{2}\left(\frac{\tau}{E}\right)^2 A_\lambda A_\lambda{}^*\sum_{k=-\infty}^{\infty} a_{mk}\zeta\sum_{k=-\infty}^{\infty}\times a_{mk}{}^*\zeta^*$$

$$\zeta=\cos\phi S(\mu-\nu+k)-i\sin\phi S'(\mu+\nu+k) \quad (37)$$

where σ^2 is the average of R deviations $\sigma_{A\mu}{}^2$ for large values of R. The averaging may be performed first for the frequencies, then for the amplitudes.

$$\sigma_A{}^2=\frac{1}{2}\left(\frac{\tau}{E}\right)^2 A_\lambda A_\lambda{}^*\sum_{k=-\infty}^{\infty} a_{mk}a_{mk}{}^*\times \int_{-\infty}^{\infty}[\cos^2\phi S^2(\mu-\nu+k)+\sin^2\phi S'^2(\mu-\nu+k)]d\mu$$

For large values of ν the terms:

$$2\frac{\sin \pi(\mu-\nu+k)}{\pi(\mu-\nu+k)}\,\frac{\sin \pi(\mu+\nu+k)}{\pi(\mu+\nu+k)}$$

of $S^2(\mu-\nu+k)$ and $S'^2(\mu-\nu+k)$ may be neglected and one obtains:

$$\sigma_\lambda^2 \doteq \frac{1}{2}\left(\frac{\tau}{E}\right)^2 A_\lambda A_\lambda^* \sum_{k=-\infty}^{\infty} a_{mk}a_{mk}^* \times$$
$$\int_{-\infty}^{\infty}\left[\left(\frac{\sin \pi(\mu-\nu+k)^2}{\pi(\mu-\nu+k)}\right)^2 + \left(\frac{\sin \pi(\mu+\nu+k)}{\pi(\mu+\nu+k)}\right)^2\right]d\mu \quad (38)$$
$$\doteq \tau^2 A_\lambda A_\lambda^*/E$$

Averaging over the amplitudes $A_\lambda A_\lambda^*$ yields the average noise power N_n in the band $-1/2 \leq \mu \leq +1/2$ or $0 \leq f \leq 1/2\tau$ in real notation.

$$\sigma^2 = \tau N_n/E = N_n/S \quad (39)$$

where $S = E/\tau$ = average signal power.

Equations 38 and 39 are the same as obtained for wire transmission of orthogonal functions.

References

1. Synchronous Communications, J. P. Costas. *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 44, 1956, p. 1713.

2. Recent Progress in Applying Information Theory to Digital Transmission Systems, R. Filipowsky. *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 77, 1958 (Jan. 1959 section), pp. 848–55.

3. Displaying Signal Spectra, A. Y. Lev, B. I. Yakhinson. *Elektrosvyas*, Moscow, USSR, no. 4, 1956.

4. Mathematical Analysis of Random Noise, S. O. Rice. "Noise and Stochastic Processes" (book). Dover Publications, Inc., New York, N. Y.

5. Communication in the Presence of Additive Gaussian Noise, R. M. Fano. "Communication Theory" (book). Academic Press Inc., New York, N. Y.

The Effect of Delay Distortion on Data Transmission

PIERRE MERTZ
NONMEMBER AIEE

DATA TRANSMISSION generally can use with economy portions of the extensive common carrier plant set up for telephony. However, this plant has been largely equalized for amplitude response and not for phase, because of the insensitivity of the ear to phase distortion.

It is important therefore to consider the effects of phase and delay distortion on the data signal. The paper starts with a brief review of the relations between phase shift and delay, and goes on to a general appraisal of possible tolerances on distortions in these characteristics.

Phase Shift and Delay in Transmission

When an unmodulated wave is transmitted over a distance its phase is shifted. The shift can be measured in degrees or radians, or also as a delay in seconds. The radian shift ϕ is equal to

$$\phi = \omega t \quad (1)$$

where

ω = radian frequency, radians/(second) = $2\pi\times$cyclic frequency, cps (cycles/sec)
t = time, sec

When an amplitude-modulated wave is transmitted, the phase of its carrier is similarly shifted. Also its envelope is shifted. Measured in seconds, this is called the "envelope delay" D. It has been shown[1] that this is equal to

$$D = d\phi/d\omega \quad (2)$$

Undistorted Transmission

When a signal is transmitted without distortion over a medium, the amplitude response of the medium is independent of frequency in the utilized band, and the phase shift is proportional to the frequency. In Fig. 1 signal B is a delayed replica of A and the amplitude response and phase shift respectively of the medium appear in dotted lines at the right.

Distorted Transmission

Where medium characteristics show a ripple or other deviation as in the solid lines at the right of Fig. 1 the signal is distorted in transmission.[2] In the case shown the distortion consists of an echo C. Often it is convenient in the phase characteristic to show only the departure from linearity as indicated in the lower right of the figure.[3]

Coarser or finer ripples in the medium characteristics indicate respectively closer-in or more remote echoes, as shown in Fig. 2.

A badly distorting circuit can be improved by the use of an "equalizer." This adds extra loss or phase shift respectively at the frequencies where these are low in the circuit, and as little as possible where they are high. Generally the further discussion in the paper holds equally for the phase distortion in the untreated circuit and for the residual distortion left after phase correction. However, usually the departures in characteristics for the untreated circuit follow broad sweeping curves; in the residuals they tend to follow fine-grained ripples.[4]

Distortion Tolerances on Data Signals

Objectives

In engineering a transmission system it is often necessary to place a tolerance limit on the distortion to be permitted in the signal. It is very simple if this tolerence can be merely set as plus and minus limits on some specific measurement on the medium. This is analogous to dimensional tolerances, say in machining a sample. It is especially desirable

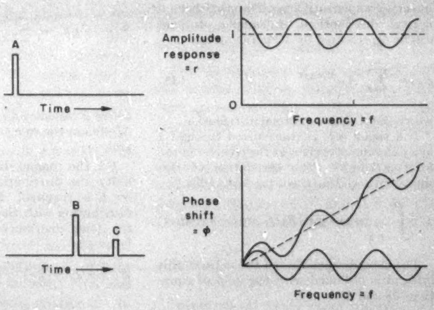


Fig. 1. Echo description in time and frequency

Paper 60-27, recommended by the AIEE Wire Communication Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., January 31–February 5, 1960. Manuscript submitted October 26, 1959; made available for printing November 13, 1959.

Pierre Mertz is with the Rand Corporation, Santa Monica, Calif.

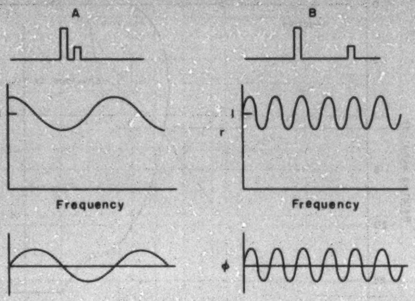


Fig. 2 (left). Comparison of close-in and remote echoes

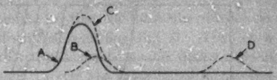


Fig. 3. Distortion from close-in and remote echoes

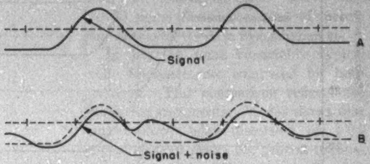


Fig. 4. Effect of noise in causing error

for this plus and minus tolerance to be independent of the exact shape of the departure. For example, in Fig. 2 it is desirable for the same tolerance to hold for coarse or fine ripples.

While there is an advantage in having this invariability exact, it is to be recognized that in a complicated system an approximation is often really sufficient. The penalty for transgressing tolerances slightly is not always catastrophic, and unless this is the case a reasonable approximation will do.

Thus the objective in appraising signal distortion is to find some measurement of the medium which is related to the penalty for the distortion in a manner that does not depend too much on the specific shape of the distortion.[5]

Echo Effects on Noise Tolerance

Fig. 3 illustrates an original signal at *A*, a close-in echo at *B*, and a remote echo at *D*. It shows that *B* distorts *A* into *C*. *C* differs from *A* chiefly by being larger, but also because the shape is changed a little. The size would be readjusted by a gain change in lining up the system; but to bring the shape back is another problem. One would guess that the close-in echo *B* would cause less penalty than the remote echo *D*. Also the effect from *B* is probably more sensitive to exact delay than that from *D*.

The nature of the penalty is suggested in Fig. 4. Here in row *B* one sees how noise can cause errors in the interpretation of a signal. Thus in the space after the first mark the added noise shown changes the space into a mark. The original signal can stand some noise before an error is caused. That is, the noise peak has to be about half the signal amplitude before an error is likely. It is clear, however, that this margin can be reduced by an echo which distorts the signal.

Thus the important penalty of a signal distortion is measured by the reduction which it causes in the margin against noise.

It is difficult to set a firm tolerance on this noise penalty because it obviously depends upon relative costs of reducing the noise and reducing the echo.[6] One may consider noise penalties of 1, 3, 6, and 10 db (decibels). A limit of 1 db is severe, to be applied in cases where it is easy to remove the echo, but expensive to eliminate the noise. A tolerance of 10 db is lax, and applies to the opposite extreme cases. A tolerance frequently used for general engineering purposes is 3 db. Of course, even if there is no noise at all, extreme distortion can cause error. This represents the maximum possible echo tolerance.[7] It is about 6 db down from the signal amplitude for the remote echoes.

A diagram which permits a simple evaluation of the noise penalty in specific cases is shown in Fig. 5(A). Here *A*, *B*, *C*, and *D* each shows three consecutive signal elements,[8] and depicts all the transitions which can occur between the middle element and the outer two. At *E* all four transitions are superimposed, together with the all mark and all space cases. The receiver sampling to detect whether the middle element is marking or spacing occurs at *F*. It is clear that a noise pulse which does not displace the signal from the upper to the lower part of the "eye" will not change a mark into a space. Similarly one which does not displace it from the lower to the upper part will not change a space into a mark.

The eye diagram with an echo present is illustrated in Fig. 5(B). This shows how the presence of the echo decreases the size of the eye, and can lead to an error with a smaller noise pulse. It also displaces slightly the best time for sampling.

By constructing diagrams such as this it is possible to plot curves of the nature shown in the solid lines of Fig. 6. These cover positive echoes i.e., where whites are echoed as whites, and blacks as blacks, in an AM (amplitude-modulation) system. In Fig. 6 the 3-db curve shows the relative echo amplitude giving a 3-db noise penalty at various echo delays measured in signal elements. The curve follows the trend which was guessed earlier. That is, larger echo

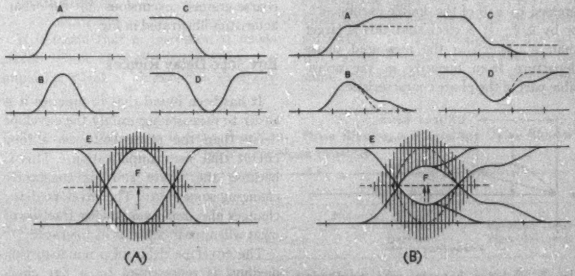


Fig. 5. "Eye" diagram

A—For undistorted signal B—With echo

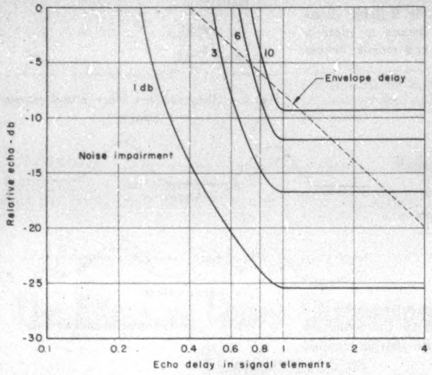


Fig. 6. Noise impairment from positive echoes of varying amplitudes and delays

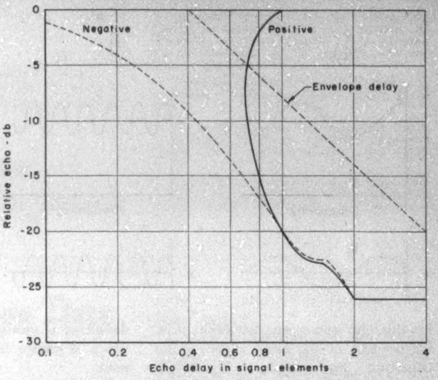


Fig. 8. Echo amplitude versus echo delay for measured departure from phase linearity of 1/10 radian

amplitudes are permissible for a close-in than for a remote echo, for a given noise penalty.

Noise impairment curves for 1, 6, and 10 db, all for positive echoes, are also plotted in Fig. 6.

ECHO VERSUS RIPPLE AMPLITUDES

In Fig. 1 it turns out that if a is the relative echo amplitude, then a is also the ripple amplitude in the response characteristic if this is plotted in radians. This is independent of the fineness of the ripple. Because of such independence it does not appear at first glance that the ripple amplitude can characterize echo penalty, which has just been shown to vary with echo delay.

However it is to be noted that when the echo delay is less than one signal element, less than one whole wave of the ripple will appear within the utilized frequency band, in the medium characteristics. This is illustrated in Fig. 7

Here the real echo amplitude is expressed by half of the double excursion h, or by $h/2$. It is clear that this is larger than h', which is the measured phase departure from linearity in the accessible part of the phase characteristic.

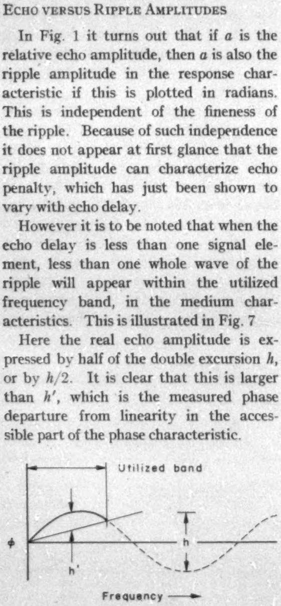


Fig. 7. Relation between echo amplitude and measured phase distortion for very-close-in echo

Thus if one sets a specific tolerance h' at all echo delays, it will correspond to larger and larger real echo amplitudes as the echoes get closer and closer in. The curve actually obtained for positive echoes is plotted in Fig. 8, for a measured h' of 1/10 radian. The curve for a negative echo is somewhat different.

SELECTION OF DISTORTION TOLERANCE

The positive curve in Fig. 8 does not really match the 1-db curve of Fig. 6, but it does show the same trend. In practice it is close enough to say that the phase departure from linearity within the utilized frequency band represents a fairly good index of the noise penalty incurred from the phase distortion, and is a good measurement on the medium upon which to place a tolerance.

It is possible in principle to set echo tolerances from the amplitude response, as well as the phase shift. In practice, however, it is found more difficult, especially say in handling the very coarse-grained excursions in the characteristics illustrated in Fig. 7.

Envelope Delay Ripples

It has been found that in practice it is easier to measure accurately the envelope delay than the phase shift, on a long circuit that uses amplification. This is because the phase shift is constantly changing somewhat. The envelope delay changes also, but by a smaller fraction of what will appear below as its tolerance.

The envelope delay of a nondistorting medium is represented by a flat characteristic, independent of frequency.

The envelope delay of the distorting medium has ripples in it which correspond in wavelength (in the frequency plot) to those in the phase shift, as shown in Fig. 9. The phase departure from linearity appears at the left, and the envelope delay at the right, for the three echoes A, B, and C.

Since the envelope delay is a derivative of the phase, the amplitude of its ripples is affected both by their fineness of structure and by the echo amplitude. In the figure the envelope delay ripples all show the same amplitude. This corresponds to a large close-in echo A, and a much smaller remote echo C. The phase departure ripples in each case are proportional in amplitude to the echoes.

ENVELOPE DELAY TOLERANCES

In the plot of Fig. 6, a fixed tolerance on envelope delay ripple corresponds to

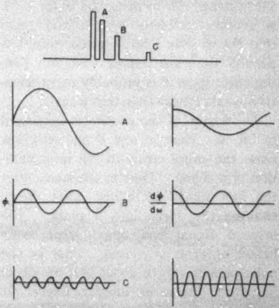


Fig. 9. Comparison between ripples in phase and in envelope delay for close-in and remote echoes

Vol.
Pt. 1
196

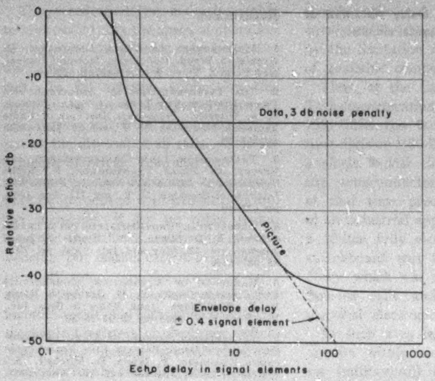


Fig. 10. Comparison of echo tolerances for data and picture transmission

a large close-in echo, but a much smaller remote echo. The course is shown by the diagonal dotted line. For fractional ripples within the utilized frequency band, i.e., at the extreme left of the plot, an additional up-turn would be experienced as discussed for the phase ripples. It is left out of the figure for simplicity.

Toward the right of the figure the fixed envelope delay tolerance for finer and finer grained ripples specifies smaller and smaller remote echo amplitudes. For echo delays of over one signal element this means smaller and smaller noise penalties. Thus a rigidly applied fixed envelope delay tolerance tends to over-engineer the system with respect to fine-grained ripples.

In practice it has been found that tolerances on the envelope delay ripples in general represent a reasonably good measure of the noise penalty, with an important proviso. The tolerance cannot be applied slavishly to fine-grained ripples, and these can be allowed larger excursions.

Two-Parameter Specfications

An envelope delay distortion characteristic represents an infinite continuum of values over some utilized frequency range. A tolerance on excursions from flatness gives a one-parameter characterization of noise penalty. It does not, however, even suggest a possible remedial or equalizing characteristic.

In unpublished memoranda of 1948, K. W. Pfleger, of Bell Telephone Laboratories, described the rather small delay distortion of wide-band filters in portions of their pass bands by two parameters. These were respectively the "slope" of the delay curve between the extremes of the utilized portion of the pass band (in the given voice channel), and the "sag" at midpoint below the straight line slope.

This description has been extended by Gibbey and Fowler[9] to cover much larger distortions occurring in voice band width transmission lines of appreciable lengths. The two-parameter description permits the advantage of a "prescription" type of equalization, i.e., based only on the parameters and not requiring a complete delay-frequency characteristic. This assumes a parabolic curve for the characteristic, as approached frequently in practice. In such cases the equalizers are chosen to have exactly the negative slope and sag, in envelope delay, of the untreated medium.

The principal weakness of the two-parameter specification, according to the authors, lies in an undue sensitivity of the noise penalty to minor variations in the exact shape of the characteristic when the penalty runs over some 3 db.

Comparison with Pictorial Transmission

It is found that in television or news-picture grade facsimile transmission the impairing effect of an echo[10] is generally similar to that for data transmission as discussed regarding Fig. 3. That is, a remote echo can be smaller than a close-in echo to give the same impairment—but after a sufficient separation the variation ceases. However, for pictorial transmission the transition comes after some 20–50 signal elements instead of merely one or two for the data transmission.

In Fig. 10 some reasonably typical echo tolerances are compared for both systems. This comparison reflects the foregoing statements. It also shows that an envelope delay distortion of ±0.4 signal element meets the general requirements for both. However, it is additionally clear that the envelope delay criterion has significance much farther into the remote echoes (or the fine structure distortion) for the pictorial than for the data transmission. By the same token the fine-structure phase departure from linearity is not as useful for the pictorial transmission.

Conclusions

It has been shown that phase or delay distortion in a propagating medium results in a noise penalty for the transmission of a signal. This penalty can be reasonably well measured in terms of the departures of the phase from linearity in the utilized frequency band; or, with precautions, in terms of the envelope delay distortion.

A tabulation, for AM transmission, of the basic tolerances on echo amplitude, phase ripples, and envelope delay ripples is given in Table I, for various noise penalties. The phase ripple magnitudes are measured in radians and degrees, and the envelope delay ripples in signal elements.

In AM systems, the figures are roughly applicable to either baseband or carrier (single or double sideband). At the very edges of the band, where signal energy is quite low, the tolerances can be relaxed somewhat. For FM (frequency-modulation) systems under approximately the same as double sideband conditions

Table I. Basic Tolerances in Echo Amplitude, Phase Ripples, and Envelope Delay Ripples for Various Noise Penalties

		Ripple Magnitude		
		Phase		
Noise Penalty, Db	Relative Echo Amplitude, Db	Radians	Degrees	Envelope Delay Signal Elements
1	−25	±0.06	± 3.5	±0.15
3	−17	±0.15	± 9	±0.4
6	−12	±0.26	±15	±0.7
10	− 9	±0.34	±20	±0.9

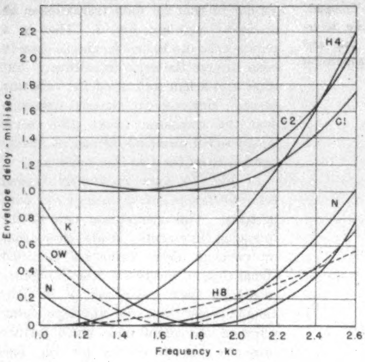


Fig. 11. Delay distortion of typical circuits

practical experience indicates the tolerances to be somewhat tighter. For FM systems where the edges of the band are appreciably cut into, i.e., leaving a total less than double sideband width, the tolerances become even tighter.

For general information the envelope delay distortions of a few typical circuits are plotted in Fig. 11. These are generally only approximate in any specific case, the exact value being subject to the precise composition of the circuit in terms of possible filters, networks, coils, wire line conductors, etc. In part they are taken from reference 9.

The circuits illustrated are:

H4 = 400 miles *H44* loaded cable
H8 = 20 miles *H88* loaded cable
K = 5 links *K* carrier
N = 4 links *N* carrier
OW = 8 repeater sections *VF* open wire
C2, *C1* = 1 link *C5* carrier, maximum and and minimum distorting channels, respectively

The *C5* carrier varies in distortion from channel to channel, mostly because of directional separating filters. The data assume no intermediate repeater; addition of these for greater lengths increases the variation among channels.

The data are plotted to displaced axes, with minimum delay equal to zero or 1 millisecond.

References

1. Measurement of Phase Distortion, **H. Nyquist, S. Brand.** *Bell System Technical Journal*, New York, N. Y., vol 9, July 1930, pp. 522–49.

2. The Interpretation of Amplitude and Phase Distortion in Terms of Paired Echoes, **H. A. Wheeler.** *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 27, June 1939, pp. 359–85.

3. Transmission Line Characteristics and Effects on Pulse Transmission, **P. Mertz.** *Symposium on Information Networks*, Polytechnic Institute of Brooklyn, N. Y., vol. III, Apr. 1954, pp. 115–30.

4. Theoretical Fundamentals on Pulse Transmission, **E. D. Sunde.** *Bell System Technical Journal*, vol 33, May and July 1954, pp. 721–88, 987–1010

5. Handbook of Automation, Computation and Control (book), **E. M. Grabbe, S. Ramo, D. E. Wooldridge.** John Wiley & Sons, Inc., New York, N. Y., 1958, pp. 18–01–18–32.

6. Transmission Aspects of Data Transmission Service Using Private Line Voice Telephone Channels, **P. Mertz, D. Mitchell.** *Bell System Technical Journal*, vol. 36, Nov. 1957, pp. 1451–86.

7. Transmission of Digital Information Over Telephone Circuits, **A. W. Horton, H. E. Vaughan.** *Ibid.*, vol. 34, May 1955, pp. 511–28.

8. Certain Topics in Telegraph Transmission Theory, **H. Nyquist.** *AIEE Transactions*, vol. 47, Apr. 1928, pp. 617–44.

9. Assessment of Effects of Delay Distortion in Data Systems, **A. D. Fowler, R. A. Gibbey.** *Ibid.*, pt. I (*Communication and Electronics*), vol. 77, 1958 (Jan. 1959 section), pp. 918–23.

10. Tolerances on the Phase Shift and Attenuation of International Television Circuits (in French), **P. Mertz.** Third Commission of Reporters, Maison des Congrès, Geneva, Switzerland, *Document no. 21*, 1947–1948, pp. 4–9.

11. Influence of Echoes on Television Transmission, **P. Mertz.** *Journal*, Society of Motion Picture and Television Engineers, New York, N. Y., vol. 60, May 1953, pp. 572–96.

12. A New Telephotograph System, **F. W. Reynolds.** *Bell System Tehnical Journal*, vol. 15, Oct. 1936, pp. 549–74.

A 240-Line Fully Electronic Telephone Switchboard

G. GOUDET
NONMEMBER AIEE

THE EVOLUTION of electromechanical switching systems shows a trend toward specialization of their functional elements and tends to their separation into two categories:

1. Those elements the function of which is to establish a connection between the calling and the called subscriber; they constitute the speech circuit.
2. Those elements the function of which is to give the instructions necessary for the operation of this circuit; they constitute the control circuit.

The latter is nothing more than a special computer and as such it must include, on the one hand, memories which contain at least the information describing the condition of the system, and on the other hand, a logical circuit which is capable, according to the previous and present data, of preparing the instructions to be given to the speech circuit. A modern switching system is therefore divided into speech circuit, logical circuit, and memories.

From the functional point of view these three parts have quite different characteristics, particularly if one compares the minimum time which each of them should devote to the communication between two subscribers. The logical circuit intervenes essentially for the establishment or the release of connections. Its action has therefore a transitory character and the duration of its interventions depends only on its own speed of operation. The situation is different with certain memories, for instance, those which register the number of a called line. This recording lasts, in fact, during the entire time necessary for dialing; it cannot be suppressed if one wants to give the logical circuit all the desirable flexibility and speed.

As for the speech circuit, it is known that, with electromechanical systems, its elements providing a path for a given conversation remain busy for its duration.

Paper 60-233, recommended by the AIEE Communication Switching Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., January 31–February 5, 1960. Manuscript submitted October 30, 1959; made available for printing December 30, 1959.

G. Goudet is with the Laboratoire Central de Télécommunications, Paris, France.

The preceding analysis makes it possible to understand that the speed of operation of electronic components may be particularly well utilized within the logical circuit. In its extreme application this speed avoids any duplication of elements, one single logical circuit being successively put at the disposal of all calls. It is the principle known as "time sharing."

Naturally, the electronic components may be introduced with benefit also into the memories, if for no other reason than for the sake of uniformity of components, but in this case it is impossible to apply completely the preceding principle. Just as a crossbar central office includes a relatively high number of electromechanical registers, an electronic switchboard must include several electronic memories for the recording of several calls simultaneously in progress.

Finally, for the speech circuit, one may envisage three different types of solutions,[1] as follows:

Semielectronic Switching System: The first type consists in keeping mechanical contacts chosen with a sufficiently high speed and low power requirement to make them operable from an electronic control circuit; one then obtains a semielectronic switching system.[2,3]

Space-Division Electronic Switching System. The second solution consists in utilizing electronic contacts according to the same schemes as those of mechanical contacts. The establishment of a connection is then performed by choosing a conducting path between the calling and called line in a complex network of contacts. Here one obtains a space-division electronic switching system.[4]

Time-Division Electronic Switching System. The third solution consists in applying to the speech circuit the time-sharing principle such as is currently used in the time-division multiplex links: samples of each conversation signal are periodically taken, the different sets of samples are interleaved, and, finally, the complex train of pulses is sent over a transmission channel. Thus, a number of simultaneous conversations can be handled on a single circuit. Following this principle one obtains a time-division electronic swiching system.[5]

Previous Development Work

It follows that several types of electronic switchboards may be conceived, differing by the extent to which the time-sharing principle is applied, from a single one-at-a-time control and a time-sharing speech circuit, to space-division speech paths and many logical control circuits. The choice is heavily influenced by the limitation in speed of operation of available components.

Thus, at the Laboratoire Central de Télécommunications, was constructed in 1951–1952 the laboratory model of a time-division 100-line switchboard with a single logical circuit. Unfortunately, the semiconductor elements available at that time precluded the feasibility of an industrial application; but in 1957 a 20-line fully electronic space-division switchboard was built for the French Navy which was capable of two simultaneous calls and which required two identical electronic registers.

In such a system no assembly need operate within times smaller than one ms (millisecond) which is very easily satisfied by the electronic components used, especially by the silicon junction diodes utilized as crosspoints in the speech circuit. This equipment has already been described.[6] It has proved its ruggedness and its perfect reliability under the very severe operating conditions met with on board a battleship. It applies only to a small extent the principle of time sharing, certain parts of the logical circuit being permanently assigned to each of the 20 lines and other parts to each of the two registers. However, this was justified by the very small capacity of the equipment.

On the contrary, the realization on the same basis of central offices up to several thousands of lines would be economically inadmissible. It would lead to too high a multiplicity of identical elements, whereas the semiconductor devices presently available make it possible to replace them by a single central logical circuit.

It is precisely to demonstrate this possibility that a 240-line PABX was built, conceived in such a way that the applied solutions can be extended to offices of up to several thousands of lines. This equipment will now be described, with the help of the general block diagram of Fig. 1 and the detailed diagrams mentioned in the course of the text.

The 240-Line PABX

SPEECH CIRCUIT

The Crosspoint. The speech circuit utilizes as an electronic crosspoint a cold cathode gas diode. This diode is similar to a tube developed at the Bell Telephone Laboratories.[7] Its firing voltage V_F is about 170 volts and its sustaining voltage V_S is about 107 volts.

This type of crosspoint is advantageous for switching because of the following property: If a set of tubes is permanently supplied at a voltage V_0 comprised between V_S and V_F, it is sufficient, for firing one of them, to apply temporarily between its terminals an additional voltage greater than $V_F - V_0$. After the suppression of this marking voltage the tube will remain fired; for this reason it is considered a crosspoint provided with a memory. This property is essential for performing the marking by the horizontal and vertical wires in a matrix of tubes. Furthermore, the gas diode used possesses a current-voltage characteristic, the slope of which is negative around the operating point, 107 volts, 11 ma (milliamperes).

The average dynamic impedance is that of a resistance of −80 ohms and of an inductance of 18 millihenrys. Advantage has been taken of the negative resistance

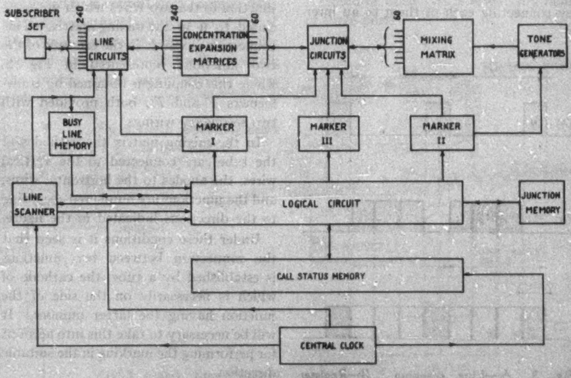


Fig. 1. General block diagram

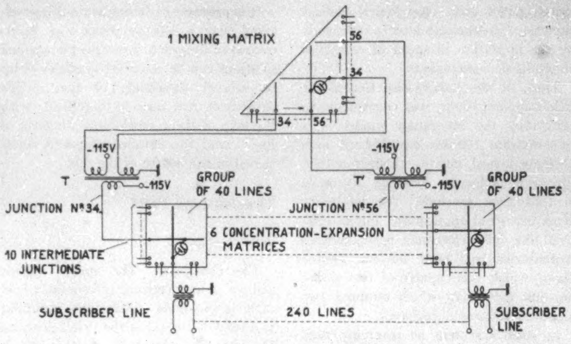


Fig. 2 Block diagram of speech circuit

for compensating the unavoidable losses in connection with the other parts of the speech circuit. At rest, the tube is equivalent to a capacity of 3 picofarads, which makes it possible to maintain cross-talk between lines at a sufficiently low level.

Structure of Speech Circuit. The speech circuit block diagram is shown in Fig. 2. It is a 1-wire circuit comprising two selection stages. In the first stage the lines are subdivided into six groups of 40 lines each. Each group of lines enters into a rectangular matrix of $40\times10=400$ gas tubes which gives access to ten outgoing intermediate junctions. There are therefore a total of 60 such junctions. Any two of these junctions can be interconnected, thanks to a triangular matrix comprising $(60\times59)/2=$ 1,770 tubes.

The communication between two subscriber lines can therefore be established by connecting each of them to an intermediate junction of its group and by interconnecting the two chosen junctions.

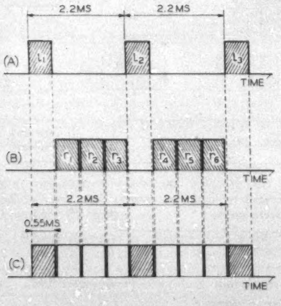


Fig. 3. A—Line scanning. B—Register scanning. C—Line and register scanning

Therefore, the first stage fulfills the functions of both concentration and expansion of the lines. The second stage fulfills the function of mixing the junctions. Each stage possesses its own d-c supply. In the matrices of concentration-expansion, the anodes of the tubes are connected to the ground through wires represented vertically in Fig. 2. The cathodes are connected through the horizontal wires to a d-c supply of -115 volts.

In the mixing matrix an intermediate junction must be connected to a horizontal and a vertical wire in order that there may be one intersection between any two junctions.

The need to give a determined bias to the horizontal wires, on one hand, and to the vertical wires, on the other, leads to utilizing a transformer for coupling a given junction to the two wires which are associated to it in the mixing matrix. This connection is visible in the example of the two junctions represented in Fig. 2, where this coupling is obtained by transformers T and T', both provided with two secondary wirings.

In the mixing matrix the cathodes of the tubes are connected to the vertical wires, the anodes to the horizontal wires, and the junctions are numbered according to the directions indicated in the figure.

Under these conditions it is seen that the connection between two junctions is established by a tube, the cathode of which is necessarily on the side of the junction having the larger number. It will be necessary to take this into account for performing the marking in the suitable direction.

With the preceding description, it may be seen that 30 simultaneous communications can be established. The possible traffic amounts to 9 erlangs with a loss probability of 1.0%.

CONTROL CIRCUIT

Main Functions. The main functions of the control circuit are:

1. Scanning of the lines to detect any new calling line.
2. For each call, recording of the called number.
3. Determination of a free path between the calling and the called lines.
4. Establishment of the connection by firing the three gas tubes chosen.
5. Supervision and release of the connection.

To these main functions must be added auxiliary functions, especially the sending in proper time of various tones: dialing tone, busy tone, ringing, and ringing back.

Scanning of Lines and of Registers. The principle of time sharing applied to a single logical circuit naturally leads it to perform a permanent scanning of the lines, one after the other, to detect a call after a waiting time which must be imperceptible for the subscriber. It may be assumed that the time necessary for the subscriber to pick up the handset is greater than 500 ms. In fact, 528 ms has been taken as the period between two scannings of a given line. The time interval comprised between two consecutive line scannings is therefore: $528/240=2.2$ ms; see Fig. 3(A). This concerns only the detection of new calls. It should be noticed that the situation is different as soon as a calling subscriber begins to dial and that it depends upon the type of signals used for this function.

Except for particular reasons resulting from the use of electronics, it has been considered undesirable to modify the conventional subscriber set. Therefore the conventional dial producing current pulses of 33 ms every 100 ms has been retained. For recording these pulses it was necessary to perform at least one line scanning within 33 ms during the dialing and to record the result obtained. For this purpose, when a line is identified as calling, it is connected to a register

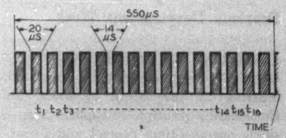


Fig. 4. Timing pulses

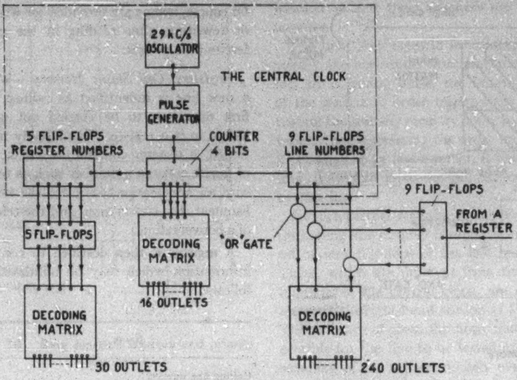

29 kC/s OSCILLATOR
THE CENTRAL CLOCK
PULSE GENERATOR
5 FLIP-FLOPS REGISTER NUMBERS
COUNTER 4 BITS
9 FLIP-FLOPS LINE NUMBERS
9 FLIP-FLOPS
FROM A REGISTER
5 FLIP-FLOPS
DECODING MATRIX
'OR' GATE
16 OUTLETS
DECODING MATRIX
30 OUTLETS
DECODING MATRIX
240 OUTLETS

Fig. 5. Central clock and its associated matrices

made of a row of ferrite cores. The register is an extremely simple device; it is not necessary to free it as soon as the dialing is finished in order to put it at the disposal of other calls. On the contrary it is advantageous to allow it to be assigned to a calling line for the whole duration of the conversation, while it gathers all the information concerning the conversation. This disposition facilitates the supervision, and especially the release, and for this reason the equipment includes a number of registers equal to the number of possible simultaneous conversations, namely, 30. The registers are permanently scanned. According to the preceding, less than $33/30=1.1$ ms should be devoted to each. Therefore, during a line-scanning period, i.e., 2.2 ms, more than two registers must be scanned. A number of three has been chosen, which leads to interleaving a line scanning and three register scannings; see Fig. 3(B). The duration of each scanning is then $2.2/4=0.55$ ms; see Fig. 3(C). During this time the logical circuit not only identifies the condition of the scanned element, but also carries out orders with a view towards making the system evolve in the desirable direction.

The following may be cited as examples: the order to look for a free junction, the order to test the called subscriber, the order to fire or to switch off a gas tube, etc.

An order is, in general, a complex operation which is subdivided into a determined sequence of elementary steps. For this reason each scanning interval is subdivided into 16 equal parts of 34 μsec (microseconds) which shall be called "time positions"; see Fig. 4. Each of these is devoted to an elementary step initiated by a 20-μsec pulse called "timing pulse" followed by a rest interval of 14 μsec.

Central Clock and Decoding Matrices: The preceding indicates that the operation of the whole control circuit is synchronized by a central clock, permanently running; see the block diagram of Fig. 5. The clock is made of a 29 kc/sec (kc per second) sine wave oscillator from which the timing pulses are derived. These pulses are recorded in binary code in a set of four flip-flops mounted as a counter. The counter is connected to two sets of flip-flops also mounted as counters. In the first one the line numbers appear successively according to the scanning rhythm indicated in Fig. 3. One line number is made of three decimal digits: the first one indicates the group number (1 to 6); the second and the third indicate the individual line number within its group (00 to 39). Each decimal digit is separately coded in binary. This results in three bits for the group number, two bits for the tens number of the line, and four bits for the units number of the line, making a total of nine bits.

In the second counter the register numbers (01 to 30) appear also in accordance with the scanning rhythm indicated in Fig. 3. Here the number is translated into binary as a whole; therefore a capacity of 5 bits is needed.

Each of the three preceding counters is connected to a diode decoding matrix which has as many outlets as elements to be counted. The designation of a line or of a register to be scanned during a given 0.55-ms interval is therefore made by marking the outlet which bears its number in the corresponding decoding matrix. The timing pulses are treated in a similar way: thanks to the decoding matrix assigned to them, each of these, $t_1, t_2, \ldots t_{16}$ is sent in proper time on an outgoing wire which is reserved for it and, through this wire, goes to a number of diode gates in the logical circuit in order to open them and to provoke the execution of an elementary step which depends on an order number indicated elsewhere.

Line Scanner. The first role of the line scanner is to examine whether a line is closed or open. It is essentially made of three parts:

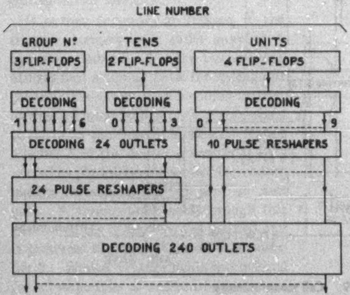

LINE NUMBER
GROUP N°
TENS
UNITS
3 FLIP-FLOPS
2 FLIP-FLOPS
4 FLIP-FLOPS
DECODING
DECODING
DECODING
1
6
0
3
0
9
DECODING 24 OUTLETS
10 PULSE RESHAPERS
24 PULSE RESHAPERS
DECODING 240 OUTLETS

Fig. 6. (left) Decoding matrix of line scanner

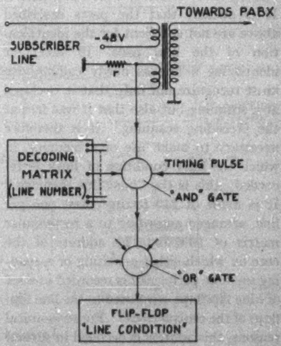

TOWARDS PABX
SUBSCRIBER LINE
-48V
r
DECODING MATRIX (LINE NUMBER)
TIMING PULSE
'AND' GATE
'OR' GATE
FLIP-FLOP 'LINE CONDITION'

Fig. 7. (right) Line circuit of line scanner

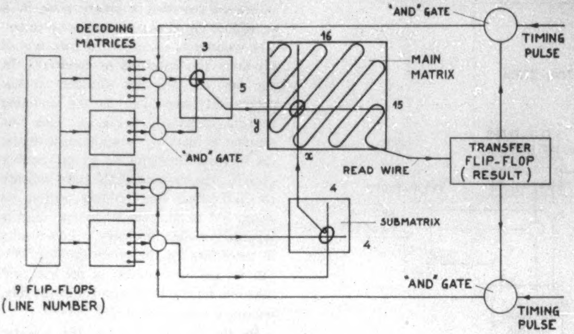


Fig. 8. Busy-line memory

1. The line number decoding matrix mentioned in the preceding paragraph.

2. The individual test circuit making it possible to pick up in each line the wanted information.

3. A flip-flop known as "line condition" in which the information CLOSED or OPEN is registered at each line scanning.

To take account of the fairly high number of lines, the decoding is made in several stages; see Fig. 6.

The individual test circuit attributed to a subscriber line is represented in Fig. 7. When the line is closed a voltage V appears across resistor r. An AND gate receives this information and the marking signal of the line, as well as a timing pulse which goes through only if the line is marked and closed. This pulse then places the flip-flop line condition in position 1. The flip-flop is common to all lines; for this reason it is driven through a set of OR gates. It is brought back to 0 by a timing pulse at the beginning of each scanning.

Busy-Line Memory, Provisional Store. It is evident that the parts described above are not sufficient for the identification of the new calls. In fact, for identifying a line as newly calling, one must recognize not only that it is closed at a scanning but also that it was free at the preceding scanning. It is therefore necessary to make use of a memory, in which the busy condition of lines is recorded; this is the "busy-line memory." It is made of 240 ferrite cores, one per line, arranged according to a rectangular matrix of 16×15. The address of the core on which either a writing or a reading must be performed is recorded in a set of nine flip-flops connected to the line flip-flops of the central clock. For economical reasons, this address is decoded in several stages comprising four diode matrices followed by two ferrite core submatrices; see Fig. 8.

During the reading operation of a core, its content is transferred to the flip-flop "result" where it is utilized. In fact, the information contained in this flip-flop is directed towards an AND gate which is also connected to the flip-flop line condition. It is only in the case of simultaneously 0 (where the line is free at the preceding scanning) in the first flip-flop, and one (closed line) in the second, that a timing pulse is permitted to go through the gate, and place it in position 1 a third flip-flop called "new calling line."

Then, if the flip-flop new calling line is in condition 1, a timing pulse produces the transfer of the line number inscribed in the input flip-flops to a new set of nine flip-flops called "provisional store," where this number remains as long as a register scanning does not indicate a free register. Of course, means are provided for writing or rewriting after reading in the main ferrite core matrix.

Registers, Call Status Memory. When a new line is indentified as calling, the first operation to be carried out is to assign to it a register. As already mentioned, a register consists only of a row of ferrite cores in a number such as to record all information useful for the establishment, the supervision, and the release of a conversation.

A register is then destined to contain information which may be tabulated as follows:

Items	Bits
Calling line number	9
Called line number	9
Units digit of junction connected to calling subscriber (number of tens not needed; in fact, each group possessing ten junctions, the tens number is identical to the group number of the calling line)	4
Units digit of junction connected to called subscriber	4
Number of order which must be carried out by the logical circuit (there is a total of 38 orders)	6
Position of duration counter utilized for temporization and for time measurements	3
Condition of line of calling or of called subscriber at time of last scanning	1
Indication that one dialing train was in course at time of the last scanning	1
Total	37

To permit possible improvements after further experience, provision is made for a capacity of 40 bits. The 30 registers therefore constitute a rectangular matrix of 30×40 elements which possess naturally the usual means of writing and reading. This assembly, known as "call status" memory, is represented in Fig. 9. It is seen that a horizontal wire goes through the 40 cores of a register; it

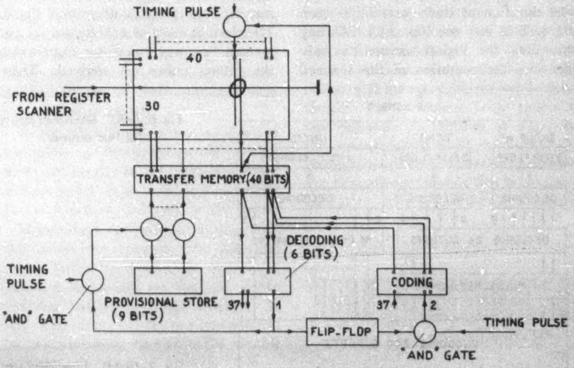


Fig. 9. Call status memory

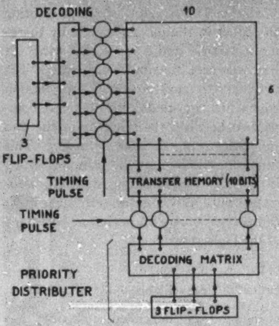


Fig. 10. Busy junction memory and priority distributor

comes from the register scanner, i.e., it is one of the 30 outlets of the decoding matrix represented in Fig. 5. A timing pulse sent on this wire at the beginning of each scanning provokes the reading which is the transfer of the information to a set of 40 flip-flops called "transfer memory," where it will be utilized. The operations to be carried out depend essentially upon the number of the order inscribed. For this reason the six flip-flops attributed to the order are connected to a 37-outlet decoding matrix (one outlet per order). In fact, there are 38 orders, but the first, called "order 0" is the order to detect a new call and, as already noted, it is executed at the beginning of each line scanning, without intervention of a register.

A free register is characterized by the presence of order number 1, called "take register." The marking voltage which then appears on the corresponding order wire causes a timing pulse to open gates which connect the provisional store (where a new calling line number is kept) to the nine first flip-flops of the transfer memory. The line number is then recorded in this memory.

It is then necessary to replace, in the transfer memory, the order number 1 by order number 2. For this purpose, one utilizes first a timing pulse for wiping out the inscribed order, then a coding matrix, having 37 inlets and 6 outlets connected to the order flip-flops, intervenes. Outlet number 1 of the decoding matrix is connected to the inlet number 2 of the coding matrix through a flip-flop and an AND gate. A timing pulse goes through this gate when the order wire 1 is marked and it produces the inscription of order number 2. Similar arrangements are utilized for all sequences of orders irrespective of whether or not their numbers are consecutive.

The next step consists in marking reversely the horizontal wire, and in opening, by a timing pulse, the input gates of the matrix in order to copy the contents of the transfer memory in the register which, then containing the order number 2 and the calling line number, is reserved from that time on to the call carried out by this line.

At each further scanning of the register, the calling line number is reproduced in a set of nine flip-flops of the line scanner. These, which are different from the line flip-flops of the central clock, are represented in the right-hand section of Fig. 5. The presence of these flip-flops makes it possible for the line to be tested at each scanning of the register namely every 22 ms.

Busy Junction Memory, Priority Distributor. The next order (number 2) is the search for a free junction within the group of the calling subscriber. It is performed by a third ferrite core memory, the "busy-junction" memory. It is made of a matrix of 6×10 cores, a row of 10 cores being attributed to each group of 10 junctions, and it possesses means of writing and reading similar to those of the preceding memories. In the block diagram represented in Fig. 10, observe the following:

1. A set of three flip-flops in which the order number 2 provokes the inscription of the group to be explored.
2. A decoding matrix which marks the corresponding row of cores.
3. A transfer memory comprising ten flip-flops.

At the beginning of the scanning, a timing pulse initiates the reading of the row marked by the decoding matrix. The content of this row then passes to the transfer memory. It is then necessary to make a choice from among the free junctions and to avoid showing a consistent preference which would make some gas tubes work more often than others.

The choice is performed by the priority distributor represented in the lower part of Fig. 10. It is a 4-bit counter which progresses by one unit at the beginning of each scanning, i.e., every 550 μsec, and contains a number comprised between 0 and 9, thus designating a preferred junction. This number is detected by a 10-outlet matrix connected to ten AND gates. These gates are also connected to the ten outlets of the transfer memory. If a junction is both preferred and free, the corresponding gate allows a timing pulse to go through and to produce the inscription of its number in the register. If the preferred junction is not free, a suitable coupling device between the flip-flops inscribes the number of the next one and so on, up to the moment where a preferred junction is found free. The junction being thus chosen, it is recorded as busy in the transfer memory; then, this information is copied in the busy-junction memory, the flip-flops of the transfer memory are brought back to 0, and finally the order number 3 is inscribed in the register.

Marker I. From that time, the register assigned to the conversation contains all information necessary for connecting the calling line to a junction by firing the gas tube located at the intersection in the concentration-expansion matrix. This is the role of marker I when it carries out order 3. The block diagram of marker I is shown in Fig. 11, where the line and the junction flip-flops of the supervision mem-

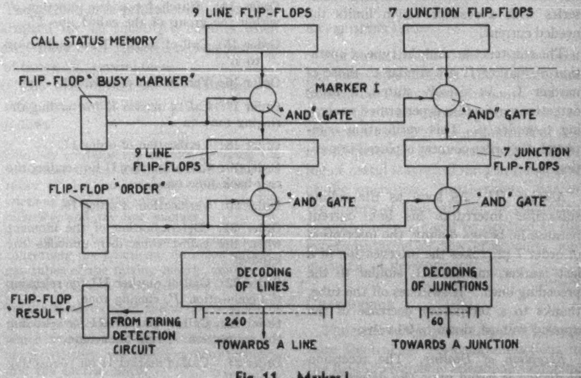


Fig. 11. Marker I

ory have been reproduced. The numbers inscribed here are first copied in the marker's own flip-flops. On account of the duration of the marking, which is normally 2 ms, the input flip-flops of the marker are protected by AND gates placed under control of the flip-flop register busy which allows the signals to go through only in case the marker is free. Then a flip-flop order intervenes, provoking the decoding of the two numbers and therefore the marking of the line and of the junction in order to fire the suitable tube. Certain precautions, described in the Appendix, have been taken in the marking process.

The result of the marking operation is verified, according to order 4, by a firing detection device, which sends an order END OF MARKING bringing back to 0 all the flip-flops of the marker. If this signal has not appeared within 20 ms, a delay device places the flip-flop result in position 0 which means that the marking has failed. One proceeds in this case to the search for a new free junction in order to perform the marking on another tube. This operation is repeated until a satisfactory result is obtained.

Marker II and Tone Generators. Marker I having carried out its work, the calling subscriber is connected to a junction and the order counter is placed in position 5. In this situation, a new marker, marker II, receives the order to connect the junction to the dialing tone generator. The four tone generators are connected to the 60 intermediate junctions thanks to an auxiliary matrix of $60\times4=240$ gas tubes. In this matrix the 60 vertical wires are those of the mixing matrix and the four horizontal wires are peculiar to the generators. To provide for the possible firing of several tubes in the same row, each tube is connected in series with a resistor which limits the needed current.

The construction and the type of operation of marker II are similar to those of marker I. As already stated, a firing detection verification is performed, according to order 6. This verification completed, the order counter is placed in position 7.

Marker III. As soon as the calling subscriber interrupts his line current because he begins dialing, the inscription of order 7 provokes the intervention of a last marker, marker III, similar to the preceding ones. It switches off the tube, thanks to a temporary decrease of the applied voltage, down to 90 volts.

Reception of Dialing. The reception of dialing is based upon the detection of the current interruptions in a line. Of course, it is necessary to compare at each register scanning, the flip-flop line condition, with the transfer flip-flop associated with the core of the register where the line condition was registered at the time of the previous scanning. The current pulses are recorded in the called subscriber flip-flops reserved for that purpose.

The order numbers 8, 9, and 10 correspond respectively to the recording of the hundreds, the tens, and the units. A duration counter is used for verifying the length of the pulses in each train as well as the time interval between two different trains. Due to this counter, release is also obtained in case of incompleted dialing.

Test of Called Subscriber. This operation is the object of order 11. The number of the called line is transferred into the input flip-flops of the busy-line memory in order to determine its idle or busy condition If it is found idle, the logical circuit records it as busy from then on and places the order counter in position 14. If it is busy the logical circuit places the order counter in position 12 and wipes the called line number of the register in order not to cut off, at the time of the release of the calling subscriber, the conversation in which the called party is engaged.

The Following Orders. The first orders have been described in detail to introduce the various constitutive parts of the PABX. For the following orders, in general, the description will be limited to citing their object without comment.

Order 12: Call of marker II in order to connect the calling junction J to the busy-tone generator.

Order 13: Verification of order 12.

Order 14: Search for a free junction J' within the group of the called line L'.

Order 15: Call of marker I for connecting J' to L'.

Order 16: Verification of order 15.

Order 17: Call of marker II for sending the ringing tone on J'.

Order 18: Verification of order 17.

Order 19: Call of marker II for sending the ring-back tone on J.

Order 20: Verification of order 19.

Order 21: Determination of the moment when the called subscriber unhooks the handset.

Order 22: Call of marker III for releasing the connection J', ringing tone.

Order 23: Call of marker III for releasing the connection J, ring-back tone.

Order 24. Call of marker II for connecting J and J'.

Marker II is, in fact, in charge of firing the tubes in the mixing matrix by marking a vertical and a horizontal wire. It has been seen that in this operation it is necessary to mark positively the junction which has the smaller number. A comparison between the numbers of the two junctions must therefore be made. This function is fulfilled by the ordinator which is a set of seven decoding matrices associated with diode gates. It makes the comparison bit by bit beginning with that of the highest rank.

Order 25. Verification of order 24.

Order 26. Surveillance of the conversation.

As soon as the called subscriber has answered, the logical circuit proceeds to the alternate surveillance of the two lines, devoting to each of them one register scanning out of two. This allows the provocation of the release by either of the two correspondents.

Orders 27–30. Operations of release, permitted by marker III.

Orders 31 and 32. Connection to the public network.

Orders 33 through 36. Repeated marking in case of failure of one tube.

Order 37: Sending of the ring-back tone to a subscriber calling the operator.

It is seen that the establishment of a conversation requires the execution of 25 orders which lasts $22\times25=550$ ms. The release is made in four orders representing a total duration of $22\times4=88$ ms.

In the preceding it will be noticed that the marking of the gas tubes has been performed stage by stage, whereas one could have imagined an end marking made over an assembly of three distinct stages. The first solution was chosen because it was the only one capable of providing for a sufficient operating reliability, taking into account the scattering of the gas tube characteristics and of their evolution in the course of time.

SUBSCRIBER'S SET

The subscriber's set, developed in cooperation with Standard Electric Lorenz, Stuttgart, Germany, involves an original ringing device. The sound signals are produced by an ordinary telephone receiver driven by a 1-transistor amplifier. They are made of 0.5-sec trains of two musical notes alternating 20 times per second. Their respective frequencies are 1,000 and 1,600 cps (cycles per second). An integrating circuit including one transistor provides for protection against line noise. Thanks to this arrangement, the ringing power has been decreased down to 1 milliwatt.

Vo
Pt.
19

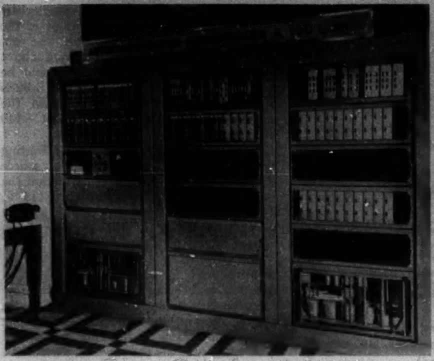


Fig. 12. Over-all view of 240-line fully electronic PABX

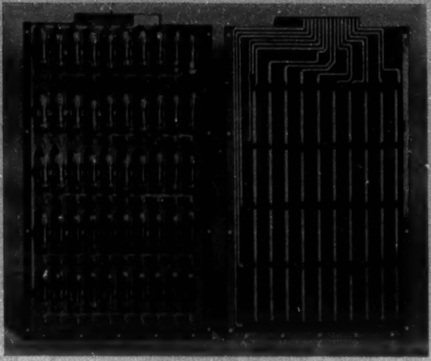


Fig. 14. Gas tube card

CONNECTION WITH PUBLIC TELEPHONE NETWORK

Three orders are devoted to the connection of subscribers to the public telephone network. This connection is not automatic. A subscriber who wants to be connected to the public network should therefore signal to the operator; he obtains this result by dialing 0. The process according to which the control circuit acts is therefore the same as in the case of an internal call, up to the dialing.

The calls coming from the exterior are directly handled by the operator, who disconnects the called subscriber from the PABX when she plugs in, and at the same time identifies him as busy. Such a process of connection with the intervention of an operator has been adopted mainly because of the experimental character of the equipment.

Fig. 13. Example of technology: boxes and cards

TECHNOLOGY

The whole circuitry is included in three adjoined bays, represented in Fig. 12. The over-all dimensions are: height 2 meters; width, 3.4 meters; depth, 0.4 meter. Each bay comprises six cases and each case contains insulating cards of approximately 20×30 centimeters, placed vertically and bearing the elementary functional circuits; see Fig. 13. On these cards the wiring is of the conventional type except for those which contain the gas tubes where the connections are made of printed circuits on both sides; see Fig. 14. The different cards are interconnected by a cabling fixed on the base, the connection between the cards and the cabling being made by plug-in connectors. The technique of printed wiring has been utilized for all repetitive circuits when they have been repeated a large number of times, i.e., several tens.

The distribution of the different circuits among the bays and cases is as follows:

One bay contains the circuits grouping the subscriber lines: the line circuits, the gas tubes of the concentration-expansion matrices as well as the marker I, the busy line memory, and the line scanner.

The second bay includes all the circuits concerning the junctions: junction circuits, gas tubes of the mixing matrix, marker II, marker III, and tone generators.

In the third bay are the main elements of the control circuit, i.e., the logical circuit which occupies a whole case, the call status memory, the busy-junctions memory, and the central clock.

Two cases in this bay are devoted to the circuits necessitated by the connection to the public network. The supplies draw their energy from a 50-cycle 3-phase power network; they take two cases respectively located in the first and in the third bay. The total power consumption is 1 kilovolt-ampere.

SOME NUMERICAL DATA

Some numerical data follow:

Over-all loss from subset to subset
  At 1,000 cps, less than 0.5 db (decibel)
  Between 300 and 3,000 cps, less than 2 db

Crosstalk level, less than −70 db

Power consumption, 1 kilovolt-ampere

Traffic with loss probability of 1%, 9 erlangs

Total number of diodes,* 7,000

Total number of transistors,* 1,500

Total number of ferrite cores,* 1,300

* Without circuits of connection to the public network.

EXPERIMENTATION

The PABX described has recently been put into operation at the Laboratoire Central de Télécommunications. Its capacity has been chosen to correspond to the present need of this organization. It is sufficiently large to permit the collecting of useful statistical data in the course of the operation. Furthermore, one of the groups of lines will not be connected to subscribers but will be devoted to the experimentation of new crosspoints now in course of development.

Conclusions

The work already done by the Laboratoire Central de Télécommunications in the field of electronic telephone switching demonstrates that it is technically possible to realize, at present, small PABXS and even large offices. It is now known that the former possess a reliability which surpasses that of electromechanical devices, in particular when severe climatic or mechanical conditions should be met. They are especially promising for military applications.

Large electronic offices are mainly destined for civil applications in which they must successfully compete with the electromechanical exchanges, i.e., with equipment which has been experienced for years and has reached a very high level of quality.

The comparison between the two kinds of solutions must be made from different angles such as: first cost, maintenance cost, reliability, flexibility, bulk and weight of equipment. Their respective capability of offering to the subscriber additional services should also be discussed.

Another important point concerns the noise generated, in particular in its relation to the problem of data transmission.

With regard to transmission, an important question concerns the methods of switching best adapted to the most modern transmission techniques such as pulse code modulation.

It must be recalled that there is no single way to electronic switching, but a variety of approaches including the semielectronic ones, which must be compared with each other. This increases tremendously the complexity of the problem of large electronic exchanges, which will not be solved without much exploratory work, theoretical as well as experimental.

It is hoped that the 240-line fully electronic PABX will be a valuable contribution to this problem, and will give the possibility of making an evaluation of great interest for the future.

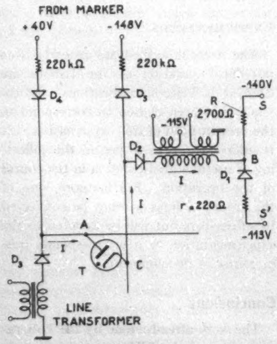


Fig. 15. Supply and marking of a tube

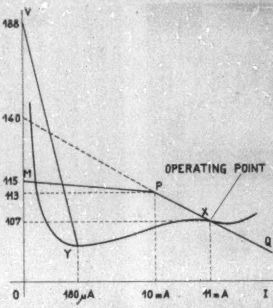


Fig. 16. Voltage-current characteristic of a gas diode (The proportions are not respected)

Appendix I. Marking Process

With a view toward understanding certain precautions taken in the marking process, it seems useful to indicate here the main numerical characteristics of the gas diode used. These are:

Firing voltage, 160 to 175 volts
Voltage applied for firing, 188 volts
Corresponding firing time, 10 ms
Sustaining voltage, 105 to 109 volts
Voltage applied for switching off, 90 volts
Corresponding switch-off time, 200 μsec

The d-c supply of the tube is described by Fig. 15 where it may be seen that the anode A is connected to ground and that the cathode C is connected indirectly to two supply sources furnishing −112 and −140 volts respectively.

To understand the roles of these two sources as well as that of resistors r, R and of diode D_1, it is sufficient to plot the curve representing the potential $-V$ of point B against current I arriving at this point. With the numerical data indicated, this curve is made of two linear segments MP and PQ. For currents I larger than 10 ma the voltage of B is superior to −113 volts and the diode D_1 is blocked. Therefore, segment PQ is located on the load line of the source of −140 volts in series with resistor R alone. On the contrary, for currents inferior to 10 ma the voltage of point B is smaller than −113 volts and the two sources supply current. Segment MP has a slope much smaller than PQ; it is approximately located on the load line of the source of −113 volts in series with resistor r alone.

For a current $I=0$, there is only a current circulating through the loop $S'BS$ and the voltage of B is $-V=-115$ volts. Therefore, when at rest the tube is submitted to a voltage $V=115$ volts.

Now if the tube is fired, thanks to a voltage temporarily larger than the firing voltage, the operating point X is located at the intersection of the V-I characteristic of the tube and of the curve made of the two preceding segments. Thus one obtains the average values indicated above ($V=107$ volts; $I=11$ ma). It should be noted in particular that when the tube is operating, diode D_1 is blocked and the supply is provided by the source S of −140 vlots only.

However, the advantage of complicating the circuit by the introduction of source S' lies in the fact that after extinction the voltage does not go up more than 115 V. There is thus a larger safety margin against spontaneous firing of tubes whose firing voltage would be abnormally low.

For performing the firing, the marker is introduced in parallel with the main supplies, care being taken to provide for the necessary decoupling, thanks to diodes D_2, D_3, D_4, and D_5. The line wire of the marker introduces a marking voltage of +40 volts, whereas the junction wire of the marker goes down to −148 volts. Therefore the marking produces temporarily a voltage of 188 volts across the tube, which primes the firing through a high resistance (440,000 ohms) to limit the current supplied by the marker. At the beginning of the marking the operating point of the tube is in Y, Fig. 16, the current supplied being only 180 per μampere. Then, in the course of a transient it moves towards the final point X already indicated. For completing the description of marker I, note lastly that the decoding of the line numbers is performed in several stages, according to Fig. 17.

The final stage, which has 240 outlets, is made of a matrix comprising diodes and resistors, and not diodes only, in view of reducing the current consumption. It can be demonstrated, in fact, that an AND gate made of two diodes is supplied at a low current only in the case when it performs the logical operation $1\times1=1$. On the contrary, an AND gate consisting of one diode and one resistor can be made so as to be supplied at a low current in all three other numerical cases of the logical multiplication. Now, in a decoding matrix, one single outlet is marked at a time. If the first type of AND gate has been used, this outlet is the only one to be supplied

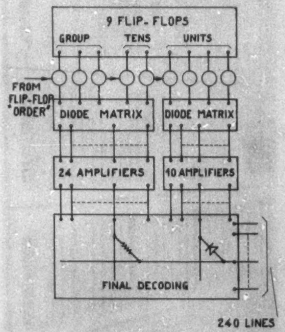


Fig. 17. Block diagram of line decoding matrix of marker I

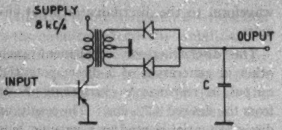


Fig. 18. Amplifier of the decoding matrices of marker I

at a low current, whereas it is the only one to be supplied at a high current if the second type has been used. With respect to current consumption, this second type is therefore much more advantageous.

The circuit of Fig. 17 involves amplifiers. They are realized according to the scheme of Fig. 18. One may remark an 8-kc power supply and the existence of a coupling transformer between the transistor and the output. This has the advantage of providing for d-c insulation and to facilitate the obtaining of fairly high output voltages with common transistors.

References

1. General Considerations on Electronic Switching, G. Goudet. *Electrical Communication*, New York, N. Y., vol. 34, June 1957, pp. 80–91.

2. An Automatic Telephone System Employing Magnetic Drum Memory, W. A. Malthaner, H. E. Vaughan. *Proceedings*, Institute of Radio Engineers, New York, N. Y., Oct. 1953, pp. 1341–47.

3. A Telephone Switching Network and Its Electronic Controls, S. T. Brewer, G. Hecht. *Bell System Technical Journal*, New York, N. Y., Mar. 1955, pp. 361–402.

4. An Experimental Switching System Using New Electronic Techniques, A. E. Joel, Jr. *Ibid.*, Sept. 1958, pp. 1091–1124.

5. Research Model for Time-Separation Integrated Communication, H. E. Vaughan. *Ibid.*, July 1959, vol. 38, pp. 909–32.

6. Fully Electronic 20-Line Automatic Telephone Exchange, C. Dumousseau. *Electrical Communication*, vol. 34, June 1957, pp. 92–101.

7. Cold Cathode Tubes for Transmission of Audio Frequency Signals, M. A. Townsend, W. A. Depp. *Bell System Technical Journal*, Nov. 1953, pp. 1371–91.

A Test Set for Measurement of Envelope Delay Distortion at Audio Frequencies with 1-Microsecond Precision

W. ALFRED CODD
MEMBER AIEE

THE accelerating use of transmission facilities for data transmission requires an increased knowledge and control of phase characteristics of the facilities and associated networks. These are commonly controlled by specification of a related quantity designated as envelope delay. Specifications have been encountered calling for control or equalization of a switching link to limits of ±10-μsec (microsecond) envelope delay over a frequency range of 1,000 to 2,600 cps (cycles per second). Such requirements call for measuring equipment capable of accurately and rapidly measuring envelope delay.

Control of phase characteristics is necessary in order to minimize distortion of signal waveforms in transmission. Distortionless transmission requires constant attenuation and constant envelope delay, resulting from insertion of a network, for all frequency components of the signal to be transmitted. If the signal in question is transmitted directly (not modulating a carrier), it also requires that the intercept of the phase characteristic on the phase axis at zero frequency be zero or an integral multiple of $\pm\pi$ radians. For a signal modulated on a carrier the latter condition is not required.

Envelope delay is equal to the derivative of the phase characteristic in radians with respect to frequency in radians per second; i.e., $d\phi/d\omega$. Audio-frequency measurements of envelope delay have been made indirectly by measuring ϕ, the insertion phase shift, at successive frequencies throughout the range of interest with sufficiently small frequency intervals so that calculations of $\Delta\phi/\Delta\omega$ from measurements at adjacent frequencies may be used as an approximation of $d\phi/d\omega$. A commercially available phasemeter which has been used for determination of envelope delay in this manner is capable of giving phase readings in a range of 30 to 20,000 cps with a precision of 0.1 degree. Changes of phase angle, at a given frequency, within a range of ±1 degree may be measured to a precision of 0.01 degree.

Determination of $d\phi/d\omega$ with the foregoing phasemeter method is very time-consuming, largely because measurement of phase at each frequency must be preceded by a multistep balancing procedure. The method has the added disadvantage that to yield a specified precision in $\Delta\omega$, it is necessary to have a higher order of precision for each test-frequency setting. Thus, a given per-cent error in the test frequencies may result in as much as 100 times this per-cent error in the resultant determination of envelope delay. This paper describes a test set which has been developed to overcome these disadvantages for envelope delay measurements in the audio-frequency range and to yield a precision of 1 μsec for measurements of envelope delay distortion up to 166 μsec.

Basic Arrangement

The test method incorporated in this test set was adopted as a direct attack on the disadvantages noted in connection with point-by-point single-frequency measurements of insertion phase shift. A means of overcoming the disadvantage arising from time-consuming rebalancing of the phase meter at each test frequency is to translate all test frequencies in a suitable manner to a single frequency at the phasemeter input, thus eliminating the need for rebalancing. Now, if a pair of test frequencies is provided which differ from each other by a suitably small increment, Δf, and arrangements are made to hold this increment constant as the test frequencies are swept through the desired test range, a basis for making the desired frequency translation is available.[1] Translation is then accomplished by passing the pair of test frequencies simultaneously through the network under test and intermodulating the pair of frequencies as they appear at the input and also at the output of the network in suitable modulator circuits. The difference frequency from each of these modulations is then used at the phasemeter inputs. Thus the desired constant frequency is provided at the phasemeter input while

---

Paper 60-46, recommended by the AIEE Data Communication Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., January 31–February 5, 1960. Manuscript submitted November 2, 1959; made available for printing February 10, 1960.

W. Alfred Codd is with the Stromberg-Carlson Company, Rochester, N. Y, a Division of General Dynamics Corporation.

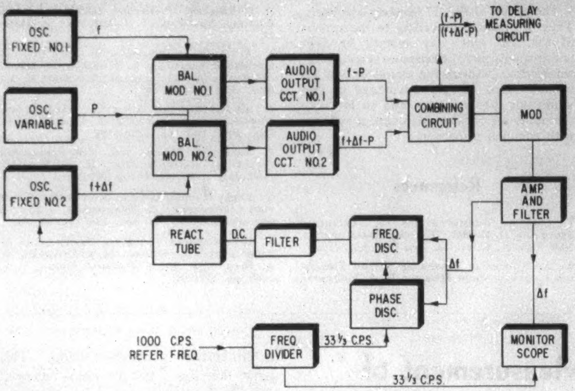


Fig. 1. **Test signal generator and control circuit**

the frequency at the test circuit is varied over the required range of measurement. With this arrangement also, the need for a high order of precision for the individual test frequencies is obviated, the requirement for precision being in maintenance of the difference between the two test frequencies which may be controlled directly.

The several components of this test set will be considered in more detail in the following sections.

Generator of Test Frequencies

The required pair of test audio frequencies is derived from three high-frequency oscillators; see Fig. 1. Two of these are "fixed" oscillators, one generating a nominal frequency of 250 kc (f) and the other a frequency, $(f+33^{1}/_{3}$ cps). The third high-frequency oscillator is variable by manual control over a range from $(f-500)$ to $(f-3{,}200)$ cps. The required audio frequencies are derived from two balanced modulators. The first fixed oscillator and the manually variable oscillator constitute the input to one modulator from which is derived an audio frequency manually variable over the range of 500 to 3,200 cps. The second fixed oscillator and the manually variable oscillator provide input to the second modulator and produce an audio frequency manually variable from $533^{1}/_{3}$ to $3{,}233^{1}/_{3}$ cps. Thus, the pair of audio frequencies is swept through the test range by the same manual control with a nominal difference between them of $33^{1}/_{3}$ cps. Although the frequency generator described provides a pair of frequencies of nominally constant difference ($33^{1}/_{3}$ cps), the difference is not maintained with adequate precision as thus generated.

The Control Circuit

The precision required for Δf is of the order of 1 in 10^4. To meet this requirement, an electric control circuit is used which measures the difference between Δf and a $33^{1}/_{3}$-cps reference frequency at any instant and provides a control voltage which is used to correct Δf by changing the frequency of the second fixed high-frequency oscillator. The arrangement of the control circuit is shown in block diagram form in Fig. 1 in relation to the generator of test frequencies.

The projected physical arrangement for the test set embodies a test signal generator assembly which incorporates the three high-frequency oscillators, the two balanced modulators and their associated audio output circuits, and the reactance-tube circuit for control of fixed oscillator no. 2. A control circuit assembly includes the other components of Fig. 1.

The combining circuit shown in this diagram is a 600-ohm resistance bridge with the two audio frequencies introduced in conjugate arms and two 600-ohm adjustable pads, feeding the control circuit modulator and the test circuit, constituting adjacent arms of the bridge.

The control circuit modulator uses a semiconductor diode to derive the difference frequency, Δf, from the pair of test frequencies for input to the discriminator elements. Two stages of amplification and associated filtering deliver this difference frequency with suitable level and waveform to the discriminator input circuits.

The discriminators supplement each other in generation of a direct-potential output in response to variations of Δf from the desired $33^{1}/_{3}$ cps. The resultant direct potential is used as a varying bias supply to the reactance-tube control element. In response to the changed bias, the reactance tube effectively alters the tuning of the tank circuit of the fixed oscillator tube thereby altering its frequency to adjust Δf in the required direction.

Initially in the design, the control circuit used a single discriminator. The phase discriminator was chosen because it has a much higher sensitivity to frequency error than has the frequency discriminator arrangement. The two circuits are basically different in principle of operation. The frequency discriminator functions by comparison of the Δf frequency input to the frequencies of resonance of a pair of tuned circuits whose geometric mean is the desired frequency to which Δf is to be controlled. It produces a direct-potential output proportional to the frequency error. It can never reduce the frequency error to zero.

The phase discriminator does not employ tuned circuits. As indicated in Fig. 1, it has two inputs, Δf and a precise $33^{1}/_{3}$-cps reference frequency derived from an accurately controlled frequency standard. It produces a direct-potential output proportional to the phase difference between Δf and the $33^{1}/_{3}$-cps reference. Considering the condition with Δf exactly equal in frequency to the $33^{1}/_{3}$-cps reference, the phase relation between the two signals is constant and the discriminator output, is supplying exactly the direct potential needed to maintain Δf at its control value. If, then, a change in Δf is somehow induced, though extremely small, its phase relation to the reference wave will change and thus produce a correcting change in direct-potential output. A stable point is reached only when the phase relation of Δf and $33^{1}/_{3}$-cps reference is static, i.e., when there is zero-frequency difference. It is seen that the phase discriminator circuit continues correction to the point of zero-frequency error. A disadvantage in use of the phase discriminator is in difficulty of adjusting for stable operation. Also, its range of control is limited from 0- to 180-degree phase difference. When the phase difference goes beyond this range, the slope of the control curve reverses and control is lost. This also permits the undesirable possibility, in initial adjustment, of lock-in of fixed oscillator no. 2 at either $(f+33^{1}/_{3})$ or $(f-33^{1}/_{3})$ cps.

Addition of the frequency discriminator to the circuit is effective in stabilizing the operation and in limiting lock-in to $(f+33^1/_3)$ cps. The frequency and phase discriminators are driven in parallel at their inputs. Their direct-potential outputs are connected in series to the bias circuit of the reactance-tube grid. This connection is through a resistance-capacitance network which serves to filter out Δf and its harmonics to prevent production of spurious frequencies in the test-frequency output by frequency modulation of the fixed oscillator. The network also serves to control the time constant of the control circuit.

The $33^1/_3$-cps reference frequency input for the phase discriminator is derived from a high-precision 1,000-cps frequency standard maintained in the testing laboratories. A 2-stage multivibrator frequency divider circuit is used to divide this down to a precise $33^1/_3$-cps output. This is in the form of short-duration pulses with $33^1/_3$-cps repetition rate.

Monitor and Adjustments for Control Circuit

The oscilloscope in Fig. 1 is provided to enable observation of the control circuit operation and facilitate adjustment to secure lock-in. The oscilloscope used is a conventional unit with horizontal and vertical deflecting systems. Input to the horizontal amplifier is the Δf signal taken from a suitable point in the control circuit amplifier. Input to the vertical amplifier is the $33^1/_3$-cps pulse signal derived from the frequency divider. When the circuit is in control with Δf equal precisely to the $33^1/_3$-cps reference, the oscilloscope pattern will show a stationary vertical pulse. When the circuit is not in control the pulse will move back and forth across the screen.

To establish control it is necessary first to adjust the input level of Δf to the frequency discriminator and also the levels of Δf and the $33^1/_3$-cps reference to the phase discriminator to specified levels. These values are not critical. In normal operation control is then established by adjustment of the voltage of a direct-potential bias in series with the discriminator output circuits until the pulse pattern on the monitor becomes stationary. This adjustment may be made with the manual frequency control of the variable oscillator set at any point on the frequency scale. With control established the dial may then be swept back and forth over the full available audio-frequency range and the difference in the pair of test frequencies delivered to the delay-measuring circuit will remain at precisely $33^1/_3$ cps.

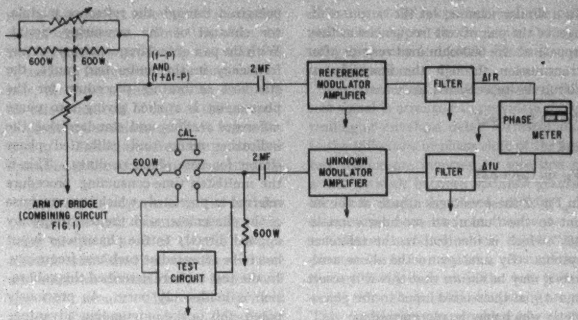


Fig. 2. Envelope delay measuring circuit

The Measuring Circuit

The pair of controlled test frequencies as delivered from the combining circuit to the measuring circuit is used as a generator voltage which is fed through a desired value of generator resistance to a specified load resistance. These are usually 600 ohms each. Switching is provided so that after initial adjustment any 4-terminal unbalanced audio-frequency network for which envelope delay measurements are desired may be inserted between the generator and load. The complete measuring circuit arrangement in block-schematic form is shown in Fig. 2. In this circuit the phasemeter is the same unit that was previously used in the indirect method of determination of envelope delay. Capabilities of the unit are those given earlier in the paper. The present arrangement as exemplified by Figs. 1 and 2 enables use of the unit to make available its incremental precision of 0.01 degree in direct measurement of envelope delay distortion as set forth in the following development.

The broad outline of the method by which this circuit provides a measurement of envelope delay has been given under the heading "Basic Arrangement." This will now be considered in more detail. Let the pair of test frequencies $(f\text{-}p)$ and $(f+\Delta f-p)$ composing the generator voltage feeding the test circuit be represented as

$$V_{11}\sin(\omega_1 t+\phi_{11})$$

and

$$V_{21}\sin(\omega_2 t+\phi_{21})$$

As indicated in Fig. 2, this voltage appears also at the input to the "reference modulator-amplifier." The basic element in this unit is a square-law modulator which, as previously indicated, derives the difference frequency, Δf_R, between the pair of test frequencies and presents this to one input circuit of the phasemeter. The circuit also includes a cathode follower input stage to provide high input impedance while permitting volume control at a relatively low impedance level to minimize effects of level adjustments on phase shift. A stage of amplification follows to provide optimum level at the modulator input. The modulator stage is a cathode-follower reflex type. A stage of amplification follows the modulator to provide a suitable level in the range required for the phasemeter input. Filtering is incorporated in and follows the modulator-amplifier to provide a pure $33^1/_3$-cps wave with necessary freedom from potentially error-producing components. Harmonics of $33^1/_3$ cps are suppressed to a level at least 40 db (decibels) below the $33^1/_3$-cps signal.

With further reference to the representation of the test frequencies, it may be shown that one product of the intermodulation process in the modulator stage can be represented as

$$kV_{11}V_{21}\sin(\omega_1 t+\phi_{11})\sin(\omega_2 t+\phi_{21})$$

By trigonometric transformation this is shown to be equal to

$$(1/2)\,kV_{11}V_{21}\cos(\omega_1 t+\phi_{11}-\omega_2 t-\phi_{21})-(1/2)kV_{11}V_{21}\cos(\omega_1 t+\phi_{11}+\omega_2 t+\phi_{21})$$

The first term of this latter expression is retained after filtering as the Δf_R output of the reference modulator–amplifier–filter system, feeding one phasemeter input. It may be written as

$$(1/2)\,kV_{11}V_{21}\cos[(\omega_1-\omega_2)t+(\phi_{11}-\phi_{21})]=(1/2)\,kV_{11}V_{21}\cos[2\pi(\Delta f_R)t+\phi_R]$$

In a similar manner, let the output voltages of the pair of test frequencies as they appear at the 600-ohm load resistor after transmission through the inserted test circuit be represented as

$$V_{12} \sin (\omega_1 t+\phi_{12})$$

and

$$V_{22} \sin (\omega_2 t+\phi_{22})$$

In Fig. 2 these voltages appear at the input to the "unknown modulator-amplifier" which is identical to the reference system. By analogy to the above analysis it may be shown that this will result in a Δf_U at the second input to the phasemeter which may be represented as

$$\begin{aligned}(1/2)\ kV_{12}V_{22} \cos\ &[(\omega_1-\omega_2)t+(\phi_{12}-\phi_{22})]\\ &=(1/2)\ kV_{12}V_{22} \cos\ [2\pi(\Delta f_U)t+\phi_U]\end{aligned}$$

Now, the functioning of the phasemeter is such that for input signals of identical frequency at its two inputs it provides a measure of the difference in the phase of the two input signals, which in the above instance will be $(\phi_R-\phi_U)$. From the foregoing development this is seen to be equal to

$$(\phi_{11}-\phi_{21})-(\phi_{12}-\phi_{22})$$

This may be rewritten as

$$(\phi_{22}-\phi_{21})-(\phi_{12}-\phi_{11})$$

From an examination of these last forms, it is seen that $(\phi_{12}-\phi_{11})$ is the phase shift incurred by f_1 as a result of insertion of the test circuit and $(\phi_{22}-\phi_{21})$ is the corresponding phase shift for f_2. Thus,

$$(\phi_R-\phi_U)=(\phi_{22}-\phi_{21})-(\phi_{12}-\phi_{11})=\Delta\phi$$

where $\Delta\phi$ is the required difference of insertion phase shift for frequencies f_1 and f_2, differing by the controlled Δf value. Since Δf is constant for all pairs of test frequencies, the phasemeter reading, $(\phi_R-\phi_U)=\Delta\phi$ in degrees, is converted to envelope delay by multiplying by a constant $=1/360\ \Delta f$. Thus, if desired, the meter may be calibrated directly in terms of envelope delay, recognizing as previously indicated that $\Delta\phi/\Delta\omega$ approximates $d\phi/d\omega$ when $\Delta\omega$ is small.

Calibration of the Measuring Circuit

Prior to use of the circuit for measurement of a test circuit, the calibrate-test key is placed in calibrate position. The phase of the signal at the input of the reference and unknown modulator channels is then identical and the phasemeter must show zero phase difference over the test-frequency range. Calibration may be accomplished by first setting up the phasemeter with both of its channels energized through the reference modulator channel of the measuring circuit. With the pair of test frequencies set at any frequency in the audio test range, the standard calibrating procedure for the phasemeter is applied giving zero phase difference reading and standardizing the indicating meter and calibrated phase shifter for true phase readings. This is the multistep time-consuming procedure referred to previously which in normal use of the phasemeter with the test frequency applied directly to the phasemeter input has to be repeated at each test frequency. In the test set here described this calibration is made only once. As previously noted, this is the outstanding advantage of this test set. It reduces testing time to the order of one tenth the time required by the measurement of phase shift at successive single frequencies, in addition to eliminating considerable amount of calculation for determination of envelope delay values.

In normal operation, when the phasemeter calibration has been completed with the pair of test frequencies at any point in the test range it is maintained over the whole range of test frequencies. This is readily checked by sweeping the manual frequency control through the range.

The next step in calibration is to return to the condition shown in Fig. 2, i.e., the reference and unknown modulator channels connected separately to the two phasemeter inputs. The phasemeter should still show zero phase difference over the test range. This may again be checked by sweeping through the test-frequency range. An adjustable decade capacitor is provided in the $33^1/_3$-cps output amplifier stage of the reference modulator-amplifier channel to use for zero adjustment when required. This completes the calibration of the test set.

Characteristics and Precision of the Measuring Circuit

The basis of the large reduction in testing time which results by the test arrangement here described has already been discussed. Another advantage also accrues directly from the translation of the variable test frequencies to a single constant frequency at the phasemeter inputs. This lies in the fact, previously noted, that the incremental technique available in the phasemeter may be used for measurement of differences in delay, thus providing a precision of 0.01 degree for measurement of changes of $\Delta\phi$ with test frequency. This may be used over a range of 2 degrees of $\Delta\phi$. With $\Delta f=33^1/_3$ cps, 0.01-degree phase difference is equivalent to 0.833, or approximately 1-μsec envelope delay. Thus, with Δf held to a precision of 1 in 10^4, envelope delay distortion up to 166 μsec may be measured with a precision of 1 μsec. When differences in phase over the frequency range are greater than 2 degrees, the incremental technique of the phasemeter cannot be applied and the precision of delay measurements is then the equivalent of 0.1 degree or 10 μsec.

It is seen that precision of 1 in 10^4 for Δf contributes, per se, negligible error in delay measurement in comparison with the figures determined by the phasemeter precision. However, this order of precision in Δf is necessary in order to maintain balance of the phasemeter because of the high envelope delay value of the internal phasemeter circuits at $33^1/_3$ cps.

To realize the potential precision given, requirements must be met at several points in the circuit relative to purity of signal in order to avoid error due to spurious signals.

It has already been noted that the harmonic content of the Δf signals at the phasemeter inputs is limited to a level lower than 40 db below the fundamental. Another potential source of error is any spurious $33^1/_3$-cps component. Assuming the extreme condition of spurious Δf being present in one phasemeter input in 90-degree phase relation to the true Δf while being absent in the other input, or in phase with true Δf, the level of the spurious component must be at least 75 db below the true component to produce no more than 0.01-degree error. Points at which such components may arise are the signal generator and the test circuit. In general, any pair of frequencies other than the test pair differing by $33^1/_3$ cps will produce a spurious $33^1/_3$-cps component by intermodulation in the measuring circuit modulators. Harmonics of the test frequencies are not a source of this type of error since they would produce harmonics of Δf when modulated. These will be eliminated by the filter circuits in the modulator channels.

It should be noted that spurious Δf components being of the same frequency as the true Δf cannot be eliminated by filtering after they have been produced and must, therefore, be controlled by control of the frequency or frequency pairs which would produce them before they arrive at the modulator stages of the measuring circuit. A means by which spurious pairs with Δf separation may arise in the signal generator is by frequency modulation of fixed oscillator no. 2 by $33^1/_3$ cps from the control circuit discriminators entering the grid circuit of the reactance tube. As previously noted, this is controlled by filtering of this component in the discriminator outputs.

The frequency of $33^{1}/_{3}$ cps should be controlled in the signal source to sufficiently low level. Unbalance in the two modulators of the signal source will produce a $33^{1}/_{3}$-cps component. This is controlled by limiting this unbalance and is further suppressed by $33^{1}/_{3}$-cps twin-T reactance-capacitance filters in the grid circuits of the modulators of the measuring circuits. Another precaution found to be desirable in the signal generator is use of d-c heater supply for the reactance tube to avoid 60-cps modulation of the oscillator encountered with 60-cps heater supply. This would produce an audio component $26^{2}/_{3}$ cps below $(f\text{-}p)$, rather than $33^{1}/_{3}$ cps. This is too close to $33^{1}/_{3}$ to be suppressed adequately by filtering and so must be kept negligible in the signal generator output.

As previously noted, spurious Δf components may also arise in the circuit under test. These also must be limited in level or suppressed by filtering out the frequency components which would give rise to them before entering the measuring circuit modulators. Note that when the two frequencies producing the spurious Δf are both frequencies other than the test pair, the level of each need be only 37.5 db below the level of each of the test pair to produce a spurious Δf 75 db below the true Δf. This follows from the square-law operation of the measuring circuit modulators.

Error which might arise in the modulator-amplifier-filter channels of the measuring circuit is avoided by the balance which is established between the reference and unknown channels before insertion of the test circuit. This was described under "Calibration of the Measuring Circuit."

Conclusions

A test set has been described which enables direct measurement of envelope delay distortion at audio frequencies with a precision of 1 μsec. This result has been obtained by generation of an audio test signal composed of a pair of audio frequencies which can be swept through the audio-frequency range with a constant difference of $33^{1}/_{3}$ cps between the pair maintained by an electric control circuit using frequency and phase discriminators in combination.

The measuring circuit employing this test signal is arranged to provide a direct reading of envelope delay corresponding to the difference in phase shift of the two test frequencies incurred by insertion of a test network between specified generator and load. Increased rapidity of measurement results from the frequency translation involved which enables the phasemeter to be used at a constant frequency ($33^{1}/_{3}$ cps) while the test pair is swept through the audio-frequency range. This translation also enables use of the incremental technique of the phasemeter employed in this set. This provides the 1-μsec precision for delay distortion up to 166 μsec. The test set has been developed in breadboard construction and has been found to perform as set forth here in measurement of several typical networks.

Reference

1. A Sweep Method for Measuring Envelope Delay, A. C. Schroeder. *LB-883*, Radio Corporation of America, RCA Laboratories Division, Industry Service Laboratory, Princeton, N. J., 12 pp.

Intermodulation Effects in FM and PM Systems

B. E. LOVE
NONMEMBER AIEE

WITH THE increasing use of radio relay systems to transmit a large variety of time and frequency division information signals, noise and signal distortions that tend to reduce the information-carrying capacity of such systems take on added importance. In any multiplex system for either speech or data-type signals, noise produced by nonlinear elements carrying the active channels generally limits the maximum signal-to-noise ratio that can be obtained. Nonlinearities in baseband amplifiers, frequency modulation generators, discriminators, and distortions of the delay characteristic are the principal contributors of this modulation noise. Thermal or first circuit noise in any part of the radio system can be readily calculated. The following discussion is concerned primarily with the measurement, evaluation, and correction of delay distortion, klystron pulling effects, and the determination of optimum system loading. A number of references are included to assist in a further study of the subject.[1–7]

Review of Repeaters

Repeaters for microwave radio systems will fall in one of two categories, either baseband or intermediate frequency. Fig. 1 illustrates the two generic types. It should be noted that the i-f (intermediate-frequency) repeater eliminates two baseband amplifiers, a discriminator, and a directly modulated klystron along with any distortion inherent in these units. However, for short haul systems, the baseband repeater equipment offers the ad-

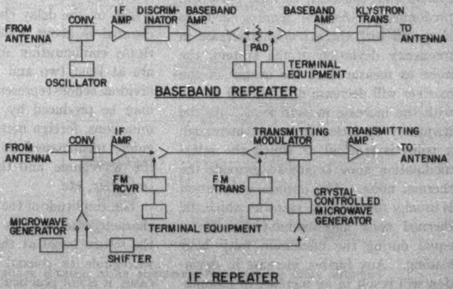


Fig. 1. Block diagram showing significant differences between repeater types

---

Paper 60-161, recommended by the AIEE Radio Communication Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., January 31–February 5, 1960. Manuscript submitted November 2, 1959; made available for printing December 9, 1959.

B. E. Love is with the American Telephone and Telegraph Company, Kansas City, Mo.

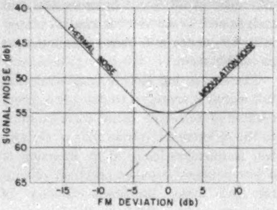


Fig. 2. Curve of system noise versus system deviation

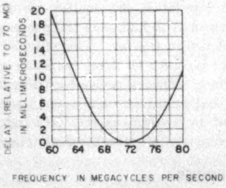


Fig. 3. Single repeater delay characteristic

vantage of easily dropping or adding information channels since the output of the receiver is at the multiplex frequencies. For systems several thousand miles in length designed for high capacity message service, it is necessary to use i-f type repeaters and equip the system with discriminators and FM (frequency-modulated) generators at message-dropping points only.

Noise Versus Loading

It will be of interest to examine the connection between thermal or first circuit noise and intermodulation noise produced by system nonlinearities. If the modulating voltage level to the FM generator is increased from some low value, and attenuators used to keep the over-all system gain constant, thermal noise will predominate. As the sending level is increased with the attendant increase in frequency deviation of the system, the noise as measured at the output of the receiver will decrease db (decibel) for db with the increase in send level. As the frequency deviation is further increased, a point is reached at which the intermodulation noise begins to override the thermal noise. The optimum deviation is usually taken as the point at which the thermal noise and modulation noise is equal during the maximum busy hour loading. Any further increase in deviation will result in an increase of modulation noise. The noise variation with FM deviation is shown graphically in Fig. 2.

Delay Characteristics

Delay characteristics may be classed into three general shapes: parabolic, sinusoidal, and slope. Parabolic shapes tend to produce third order distortion products while slope produces second order products. The distortion products resulting from sinusoidal shapes depend upon the ripple period and amplitude as well as the phase with respect to the frequency position of the unmodulated carrier. The shapes found in practice will generally be some combination of the three basic shapes.

Ideally, the delay in a channel should be constant with frequency. The absolute delay is usually of little consequence. A typical delay characteristic of a repeater is shown in Fig. 3. The vertical scale here is such that 10 millimicroseconds is equivalent to one degree. Since this shape is quite consistent between repeaters, it is readily reduced by providing an equalizer with a complementary shape in each repeater. Variations due to the *Q* of cavities, broadness of filters, etc., at different frequencies can be controlled by adding or removing equalizers. As an example, Fig. 4 shows an overequalized channel about 500 miles long. It was necessary to remove the parabolic equalizers from about one third of the repeaters for the best equalization.

Figs. 5(A) and (B) show delay characteristics having a general long-period sinusoidal shape with some residual slope. These irregular shapes were caused by mismatched impedances in the i-f circuits at a number of repeaters. The signal-to-noise ratio on these channels was about 30 db during busy periods. The objective would be in the order of 55 db.

Fig. 6 shows the effects of waveguide echoes on the delay characteristic. This channel is about 250 miles long. The ripple configuration suggests that there are at least two and possibly three individual echoes represented. Such echoes may be produced by poor antenna terminations, foreign material in the waveguide, mismatched impedances between the waveguide and the receiver crystal converter, etc.

The amplitude of the ripple in the delay characteristic depends upon the sum of the return losses at the two ends of the waveguide or coaxial cable. In many cases, it is not practicable to improve the

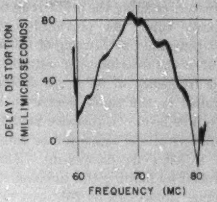


Fig. 4. Measured characteristic of over-equalized channel

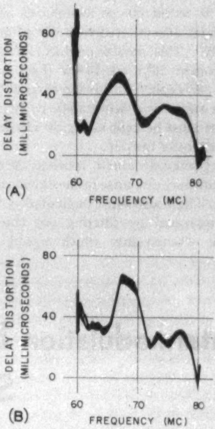


Fig. 5. Delay distortions caused by mismatched i-f impedances

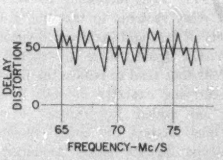


Fig. 6. Delay distortion from waveguide echoes

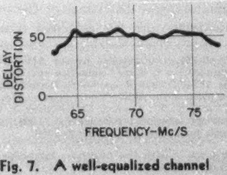


Fig. 7. A well-equalized channel

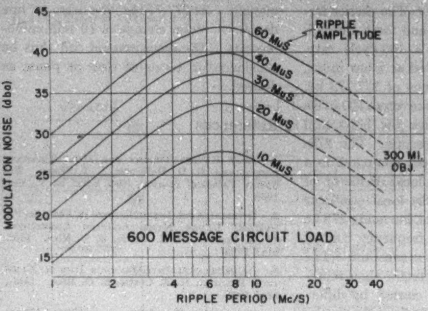


Fig. 8. Curves showing the telephone circuit noise caused by delay ripples of various periods and amplitudes

return loss by obtaining better impedance matches because of design or other limitations. An example is the feed for a parabolic antenna. Another example is the receiver crystal converter. Here the impedance looking into the crystal cavity is dependent upon the individual crystal characteristics as well as the current through them.

Ferrite devices, such as isolators, are very useful in either improving impedance matches or improving the operation of some units of equipment. As an example, an isolator may be used at a crystal converter to improve the return loss as seen by the waveguide. Since the isolator has little attenuation in the forward direction and high attenuation in the backward direction, the return loss from the waveguide side will be improved by some amount.

The impedance reflected by an antenna into the frequency-determining cavity of a directly modulated klystron oscillator often produces a frequency deviation that is not solely a function of the modulating voltage. This effect is commonly termed klystron pulling. An isolator placed very close to the klystron is very effective in controlling this condition. It is now common practice to design the microwave equipment using directly modulated klystrons to include an isolator.

Some systems exhibit a delay characteristic that has a slope either increasing or decreasing with frequency. This shape is readily controlled by appropriate slope equalizers. The principal control over sinusoidal variations in the delay characteristic is by control of the impedance matches as implied in the discussion above. This control can be exercised by appropriate maintenance procedures and the provision of the necessary equipment such as isolators. The characteristic of a well-equalized channel is shown in Fig. 7.

Modulation Noise Versus Delay Distortion

The modulation noise resulting from sinusoidal variations in the delay characteristic can be calculated with a fair degree of accuracy. The curves shown in Fig. 8 were calculated for a number of different amplitudes and noise plotted against the ripple period for a 600-message circuit load. Order of magnitude estimates of the noise produced by a complex characteristic can be obtained by breaking it up into component sinusoids of constant amplitude at the various periods and determining the resultant noise contribution of each. These may then be added on a power basis to obtain the final result. The broad peak of each of the curves in the neighborhood of 6 to 8 mc (megacycles per second) is quite apparent. The peaks occur at a ripple period that is a little more than twice the frequency of the top message channel. Some characteristics, as seen in Figs. 5 and 6, have been found with ripple periods in the vicinity of 7 mc. The message circuit noise as measured on these channels demonstrated the existence of the broad peaks.

Evaluation Methods

The curves of Fig. 8 show the modulation noise resulting from delay distortion having a particular shape. Even with an exact knowledge of the noise contributed by delay irregularities, no information is conveyed about the contributions from nonlinearities in baseband amplifiers, FM generators, and discriminators. Some method of evaluating the over-all system performance is therefore desirable. Two methods have been used. The first method uses a broad-band source of white noise to simulate the desired number of active message channels. The other uses the common 2-tone intermodulation test procedure.

The broad-band noise source is first filtered to restrict the bandwidth to that occupied by the multiplex system. Two slots, each about 3 kc wide, are made free of noise by appropriate rejection filters. Frequency-wise, the slots are located near the extremes of the noise spectrum. This filtered noise source is then used to modulate the system. At the receiving end of the system, some type of selective power-measuring equipment must be provided to measure the modulation products that now are found in the slots that were cleaned out at the transmitting end. The two slots are equivalent to two idle message channels. The noise power measured in them is the summation of all possible combinations of modulation products occurring as a result of system nonlinearities acting on the noise-modulated signal.

The intermodulation method makes use

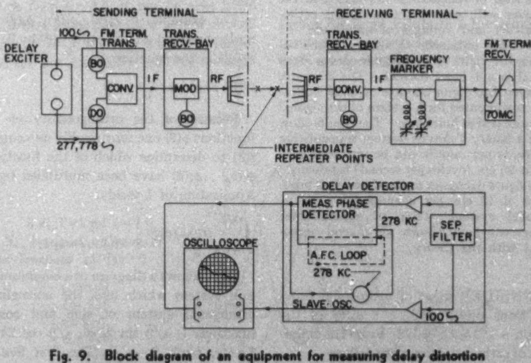


Fig. 9. Block diagram of an equipment for measuring delay distortion

of two tones having frequencies separated by perhaps 100 kc and having equal amplitudes. The system is then modulated with the two tones and suitable measuring equipment provided to determine the level of the second and third order modulation products. Measurements should be made using several pairs of tone frequencies which are not harmonically related.

With either method, the deviation is increased in small increments, say 2 db, from some small value considerably below the maximum rated deviation of the equipment. By keeping the system gain constant with attenuators, a plot of the results will yield results similar to Fig. 2. The results using the noise method may be applied almost directly to message systems while the intermodulation method requires the development of an empirical relationship for such application.

Delay Distortion Measurement

The block schematic of one type of equipment for measuring delay distortion is shown in Fig. 9. The carrier frequency is swept back and forth through the i-f pass band at a 100-cycle rate. This carrier is modulated at a low index, about 0.3, at a frequency of 277.778 kc. At the receiver, the recovered 100-cycle voltage is used to drive the horizontal deflection of an oscilloscope. The 278-kc voltage is continuously compared in a phase detector with a local oscillator of similar frequency. This local oscillator is locked in step with the incoming 278 kc through an automatic frequency control circuit having a long time constant. The output of the phase detector then represents the phase error, caused by differences in delay, continuously as the carrier is moved from one edge of the pass band to the other. The voltage representing the phase error is applied to the vertical deflection plates of the oscilloscope. Marker pips may be introduced by absorption wavemeters for the purpose of determining band edges, ripple periods, etc. The oscilloscope presentation then represents the delay variations across the band in terms of phase error versus frequency. At 277.778 kc, the delay for one degree of phase change is 10 millimicroseconds. The measurement can thus be considered in terms of time or phase as desired.

References

1. LOAD RATING THEORY FOR MULTI-CHANNEL AMPLIFIERS, B. D. Holbrook, J. T. Dixon. *Bell System Technical Journal*, New York, N. Y., Oct. 1939.

2. INTERCHANNEL INTERFERENCE IN FM AND PM SYSTEMS UNDER NOISE LOADING CONDITIONS, W. R. Bennett, H. E. Curtis, S. O. Rice. *Ibid.*, May 1955.

3. INTERCHANNEL INTERFERENCE DUE TO KLYSTRON PULLING, H. E. Curtis, S. O. Rice. *Ibid.*, May 1957.

4. DISTORTION PRODUCED IN A NOISE MODULATED FM SIGNAL BY NON-LINEAR ATTENUATION AND PHASE SHIFT, S. O. Rice. *Ibid.*, July 1957.

5. EXPERIMENTAL TRANSVERSAL EQUALIZER FOR TD-2 RADIO RELAY SYSTEMS, B. C. Bellows, R. S. Graham. *Ibid.*, Nov. 1957.

6. DISTORTION IN FM SYSTEMS DUE TO SMALL SINUSOIDAL VARIATIONS OF TRANSMISSION CHARACTERISTICS, R. G. Medhurst, F. F. Small. *Proceedings*, Institute of Radio Engineers, New York, N. Y., Nov. 1957.

7. REDUCTION OF INTERMODULATION IN MICROWAVE SYSTEMS BY USING FERRITE ISOLATORS, N. P. Weinhouse. *National Convention Record* Institute of Radio Engineers, pt. 8, 1958.

On the Transmission of Information by Orthogonal Time Functions

H. F. HARMUTH
ASSOCIATE MEMBER AIEE

Synopsis: In a digital communication system each digit of a character may be represented by one function of an orthogonal set of functions. This function is multiplied by +1 or −1 if binary digits are being used. All functions of the character are added and transmitted in parallel. This sum of functions is correlated in the receiver with each function of the orthogonal set and voltages of amplitude +1 or −1 are obtained. No interference between the digits of a character occurs if the signal is not band limited. The crosstalk between the digits because of band limitation is computed for a system of orthogonal sine and cosine functions. The results show that a binary Teletype system transmitting 60 words per minute can be operated with about 25 cps (cycles per second) bandwidth. This figure decreases to the theoretical minimum of 15 cps per channel for a multichannel system. Tests of experimental equipment have shown very good agreement with the theory.

CONSIDER a set of orthogonal functions $\phi_1(\theta)$, $\phi_2(\theta)$, $\ldots$ $\phi_l(\theta)$ in the region $-1/2 \leqslant \theta \leqslant +1/2$. From the definition of orthogonal functions follows

$$\int_{-1/2}^{1/2} \phi_l(\theta)\phi_m(\theta)d\theta \begin{cases} =K \text{ for } l=m \\ =0 \text{ for } l\neq m \end{cases} \qquad (1)$$

K is assumed 0 normalized to 1.

If the functions $\phi(\theta)$ are multiplied by 1 or 0 and added, one obtains a function $f(\theta)$.

For example, if $\phi_1(\theta)$, $\phi_2(\theta)$, $\phi_3(\theta)$ are multiplied by 1 and all other $\phi(\theta)$ by 0, the result is the function

$$f(\theta)=\phi_1(\theta)+\phi_2(\theta)+\phi_3(\theta) \qquad (2)$$

Because of the orthogonality of the functions $\phi(\theta)$ one may readily decompose $f(\theta)$ to determine which of the functions $\phi_1(\theta)\ldots\phi_l(\theta)$ have been multiplied by 1. Application of 1 yields

$$\int_{-1/2}^{1/2} f(\theta)\phi_l(\theta)d\theta \begin{cases} =1 \text{ for } l=1,\,2,\,3 \\ =0 \text{ for } l=4,\,5,\,\ldots \end{cases}$$

Fig. 1 shows a diagram of a communication system which uses the normalized orthogonal system of sine and cosine functions 1, $\sqrt{2}\sin 2k\pi\theta$, $\sqrt{2}\cos 2k\pi\theta$; $\theta=t/\tau$, $k=1,\,2,\,\ldots$. The first five of these functions are shown in Fig. 2. Assume that the five digit character 11001 is to be transmitted. The switches (multipliers) in the transmitter let through 1, $\sqrt{2}\sin 2\pi\theta$, and $\sqrt{2}\cos 4\pi\theta$, but not $\sqrt{2}\cos 2\pi\theta$ and $\sqrt{2}\sin 4\pi\theta$. The transformers add the voltages and yield

$$f(\theta)=1+\sqrt{2}\sin 2\pi\theta+\sqrt{2}\cos 4\pi\theta \qquad (3)$$

The function $f(\theta)$ varies continuously with time and is similar to a voice signal. It may be transmitted by amplitude or frequency modulation, or any other type of modulation suitable for the transmission of continuously varing time functions.

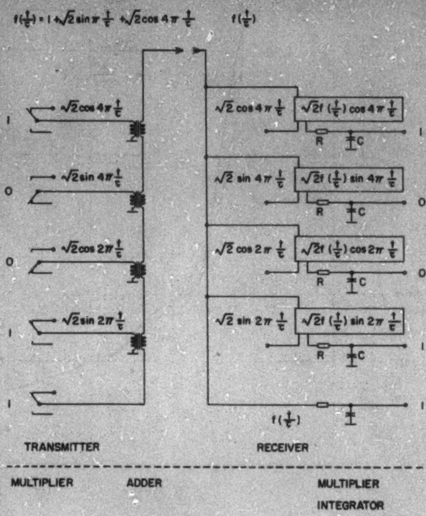
In the receiver $f(\theta)$ is multiplied by the same five functions which were used in the transmitter, the integral over the period τ is taken and the character 11001 multiplied by τ is recovered

$$\tau\int_{-1/2}^{1/2} f(\theta)(1)d\theta=\tau$$

Paper 59-1076, recommended by the AIEE Communication Theory Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting Chicago, Ill., October 11–16, 1959. Manuscript submitted May 25, 1959; made available for printing December 17, 1959.

H. F. HARMUTH is with the Research Division of the Stromberg-Carlson Company, Rochester, N. Y., a Division of the General Dynamics Corporation.

The author wishes to express his gratitude to Mr. R. C. Clark, Mr. W. Brouillette, Mr. C. Nielsen, and Mr. H. Putschi of the General Electric Company, Electronics Laboratory, Syracuse, N. Y., for several helpful discussions on the subject.



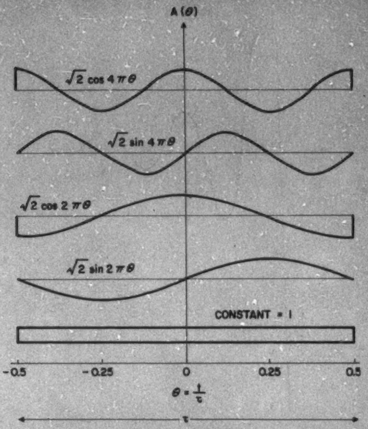


Fig. 1 (left). Communication system using a set of orthogonal trigonometric functions

Fig. 2 (above). Set of orthogonal trigonometric functions

$$\tau\int_{-1/2}^{1/2} f(\theta)\sqrt{2}\sin 2\pi\theta d\theta=\tau$$

$$\tau\int_{-1/2}^{1/2} f(\theta)\sqrt{2}\cos 2\pi\theta d\theta=0$$

$$\tau\int_{-1/2}^{1/2} f(\theta)\sqrt{2}\sin 4\pi\theta d\theta=0$$

$$\tau\int_{-1/2}^{1/2} f(\theta)\sqrt{2}\cos 4\pi\theta d\theta=\tau$$

In the following only trigonometric functions are considered but there is no difficulty other then computational in applying other sets of orthogonal functions.

Transmission Through a Noise-Free Channel

The bandwidth required for the communication system may be derived as follows: $f(\theta)$ is the sum of several sine and cosine terms; hence its frequency spectrum is the sum of the frequency spectra of the sine and cosine terms. The frequency spectra a ν for 1, $\sqrt{2}$ sin $2k\pi\theta$, and $\sqrt{2}$ cos $2k\pi\theta$, $(-1/2\leqslant\theta\leqslant+1/2)$, may be readily computed

$$a(\nu_1 1)=\int_{-1/2}^{1/2}\cos 2\pi\nu\theta d\theta=\sin\pi\nu/\pi\nu \qquad (4)$$

$$a(\nu_1\sqrt{2}\sin 2k\pi\theta)=$$
$$\frac{1}{\sqrt{2}}\left(\frac{\sin\pi(\nu-k)}{\pi(\nu-k)}-\frac{\sin\pi(\nu+k)}{\pi(\nu+k)}\right) \qquad (5)$$

$$a(\nu_1\sqrt{2}\cos 2k\pi\theta)=$$
$$=\frac{1}{\sqrt{2}}\left(\frac{\sin\pi(\nu-k)}{\pi(\nu-k)}\right)+\frac{\sin\pi(\nu+k)}{\pi(\nu+k)} \qquad (6)$$

Fig. 3 shows the frequency spectra of the first five orthogonal trigonometric functions. It can be seen that most of the energy is concentrated at low frequencies. For the comparison with standard PCM systems the dot frequency f_d is introduced

$$f_d=1/2T=n/2\tau \qquad (7)$$

T is the average time available for the transmission of one digit, τ the average time available for the transmission of one character, and n the number of digits in one character. For five digit characters follows $f_d\tau=\nu_d=2.5$, it can be seen from Fig. 3 that a low pass filter with a cut-off frequency $\nu=3$, which is only 20% larger then the dot frequency, will let through almost all energy. A communication system using n-digit characters requires for $n=2l+1$ or $n=2l$ the bandwidth

$$w\tau=\nu=l+\frac{1}{2}+\alpha=\nu_d+\alpha \qquad (8)$$

α is a constant determined by the tolerable distortion of the characters because of band limitation; in the foregoing example α was chosen equal to 1/2. Introduction of equation 7 into equation 8 yields

$$w=\left(1-\frac{\alpha}{l+1/2}\right)/2T \quad \text{for } n=2l+1 \qquad (9)$$

$$w=\left(1+\frac{\alpha+1/2}{l}\right)/2T \quad \text{for } n=2l \qquad (10)$$

Hence for large values of l the required bandwidth approaches its theoretical limit $w=f_d=1/2T$.

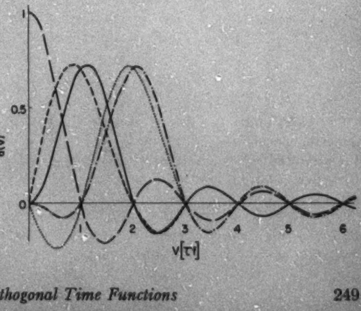


Fig. 3. Frequency spectra of the functions of Fig. 2

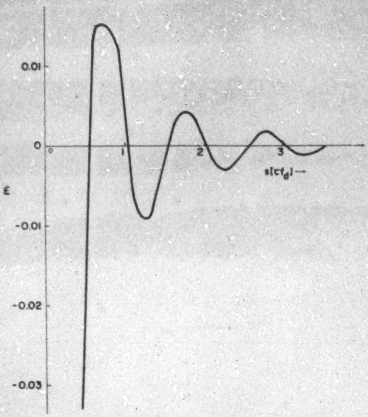


Fig. 4 (left). Error versus cutoff frequency for the constant term, $E=2\pi^{-1}(\sin 2\pi s+\cos 2\pi s/2\pi s)$. The ideal limit of the bandwidth corresponds to $s=1/2$

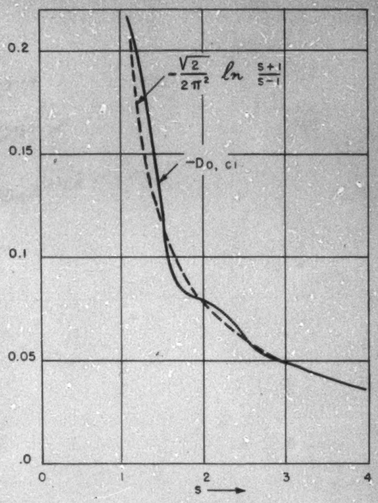


Fig. 5 (right) Crosstalk from the digits transmitted by $\sqrt{2}\cos 2l\pi\theta$ to the digit transmitted by the constant voltage

$D_{o,ci}=-1/2\sqrt{2}\,\pi^{-2}\{-\mathrm{ci}\ 2\pi(s-1)+\mathrm{ci}\ 2\pi(s+1)-\ln[(s+1)/(s-1)]\}$

The value of α may be determined by cutting off all frequencies $\nu>s$ and investigating the resulting distortions of $f(\theta)$. The assumption of an ideal low-pass filter is only made to illustrate the effects of band limitation. The filter is not used to suppress noise. It is possible to study the distortions of the individual sine and cosine functions which make up $f(\theta)$. The inverse Fourier integral taken over the frequency spectrum equation 4 yields, if no frequencies higher then s are passed

$$F_0(\theta)=\int_{-s}^{s}\frac{\sin\pi\nu}{\pi\nu}\cos 2\pi\nu\theta d\nu$$
$$=\frac{1}{\pi}[\mathrm{Si}\ \pi s(2\theta+1)-\mathrm{Si}\ \pi s(2\theta-1)] \quad (11)$$

where Si x is the sine integral

For the sine and cosine terms equations 5, 6 one obtains

$$F_{s,k}(\theta)=\frac{\sin 2k\pi\theta}{\sqrt{2}\pi}[\mathrm{Si}\ \pi(s-k)(2\theta+1)-$$
$$\mathrm{Si}\ \pi(s-k)(2\theta-1)+\mathrm{Si}\ \pi(s+k)(2\theta+1)-$$
$$\mathrm{Si}\ \pi(s+k)(2\theta-1)] \quad (12)$$

$$F_{c,k}(\theta)=\frac{\cos 2k\pi\theta}{\sqrt{2}\pi}[\mathrm{Si}\ \pi(s-k)(2\theta+1)-$$
$$\mathrm{Si}\ \pi(s-k)(2\theta-1)+\mathrm{Si}\ \pi(s+k)(2\theta+1)-$$
$$\mathrm{Si}\ \pi(s+k)(2\theta-1)] \quad (13)$$

$$\mathrm{Si}\ x=\frac{\pi}{2}+\mathrm{si}\ x=\int_0^x\frac{\sin y}{y}dy$$

The band-limited function $f(\theta)$ is represented as sum of $F_0(\theta)$, $F_{s,k}(\theta)$, and $F_{c,k}(\theta)$. For instance the $f(\theta)$ corresponding to the character 11001 has now the form

$$f(\theta)=F_0(\theta)+F_{s,1}(\theta)+F_{c,2}(\theta) \quad (14)$$

instead of the form of equation 3. The distortion of the received character may be obtained by multiplying equations 11, 12, 13 with the various sine and cosine functions and integrating over the period τ. Two kinds of integrals have to be distinguished: First, the "information integrals" with a value close to 1

$$\int_{-1/2}^{1/2}F_0(\theta)d\theta;\ \int_{-1/2}^{1/2}F_{s,k}(\theta)\sqrt{2}\sin 2k\pi\theta d\theta;$$
$$\int_{-1/2}^{1/2}F_{c,k}(\theta)\sqrt{2}\cos 2k\pi\theta d\theta \quad (15)$$

Second, the "crosstalk" integrals, which will be 0, unless the frequency limitation produces some crosstalk between the digits of a character

$$\int_{-1/2}^{1/2}F_{s,k}(\theta)d\theta;\ \int_{-1/2}^{1/2}F_{s,k}(\theta)\sqrt{2}\times$$
$$\sin 2m\pi d;$$
$$\int_{-1/2}^{1/2}F_{s,k}(\theta)\sqrt{2}\cos 2l\pi\theta d\theta \quad (16)$$

$$\int_{-1/2}^{1/2}F_{c,k}(\theta)d\theta;\ \int_{-1/2}^{1/2}F_{c,k}(\theta)\sin 2l\pi\theta d\theta;$$
$$\int_{-1/2}^{1/2}F_{c,k}(\theta)\sqrt{2}\cos 2m\pi\theta d\theta$$

$$\int_{-1/2}^{1/2}F_0(\theta)\sqrt{2}\sin 2l\pi\theta d\theta;$$
$$\int_{-1/2}^{1/2}F_0(\theta)\sqrt{2}\cos 2l\pi\theta d\theta;$$
$$k,\ l,\ m=1,\ 2,\ 3\ldots \quad k\neq m$$

The first integral of equation 15 yields by partial integration

$$\int_{-1/2}^{1/2}F_0(\theta)d\theta=1+\frac{2}{\pi}\left(\mathrm{si}\ 2\pi s+\frac{\cos 2\pi s}{2\pi s}\right)$$
$$=1+E \quad (17)$$

The second term in equation 17 yields the distortion of the received digit because of band limitation. Using the asymptotic expansion

$$\mathrm{si}\ x\approx -0!\frac{\cos x}{x}-1!\frac{\sin x}{x^2}+2!\frac{\cos x}{x^3}+$$
$$3!\frac{\sin x}{x^4} \quad (18)$$

E may be written in the following form

$$E\approx-\frac{2}{\pi}\left[\frac{\sin 2\pi s}{(2\pi s)^2}-2\frac{\cos 2\pi s}{(2\pi s)^3}\right] \quad (19)$$

If α is chosen equal 1/2 in equation 9, it follows $s=w\tau=1$ and therefore

$$E\doteq\frac{2}{\pi}\frac{2}{(2\pi^3)}\doteq 0.005$$

Hence the error is only one-half per cent. E is plotted versus $s=1/2+\alpha$ in Fig. 4, according to equation 17.

The second and third integral of equation 15 yield

$$\int_{-1/2}^{1/2}F_{s,k}(\theta)\sqrt{2}\sin 2k\pi\theta d\theta$$
$$=\int_{-1/2}^{1/2}F_{c,k}(\theta)\sqrt{2}\cos 2k\pi\theta d\theta=1+E$$
$$(20)$$

$$E=\frac{1}{\pi}\left[\mathrm{si}\ 2\pi(s-k)+\frac{\cos 2\pi(s-k)}{2\pi(s-k}+\right.$$
$$\mathrm{si}\ 2\pi(s+k)+\frac{\cos 2\pi(s+k)}{2\pi(s+k)} \quad (21)$$

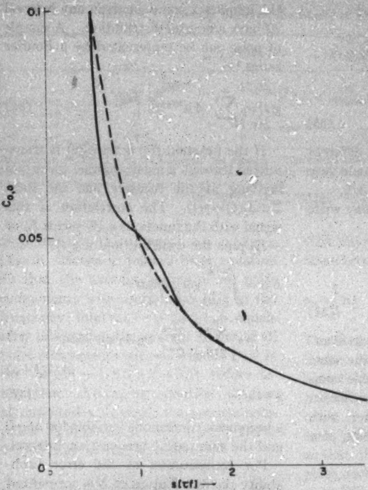


Fig. 6 (left). Crosstalk between the constant terms of two adjacent characters. The ideal limit of the bandwidth corresponds to s=1/2

- - - $1/2\pi^2 s$

—— $C_{0,0}=(2\pi^2 s)^{-1}\ (1+\sin 2\pi s - 1/2\ .\ \sin 4\pi s/4\pi s)$

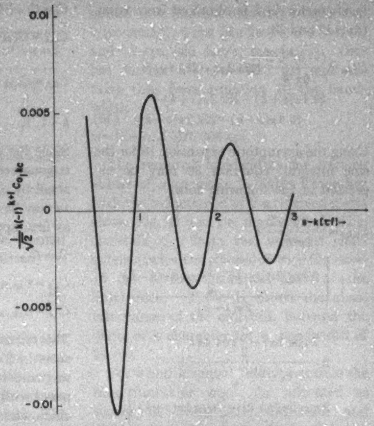


Fig. 7 (right). Crosstalk between the constant term of character l and the cosine terms of character l+1. The ideal limit of the bandwidth corresponds to s=1/2

$2^{-1/2}k(-1)^{k+1}C_{0,kc}=\sin\ 2\pi(s-k)/2\pi(s-k)\quad [2+\cos\ 2\pi(s-k)]$

In first approximation E is represented by the graph of Fig. 4 if $s-k$ is substituted for s.

The interference between the digits of one character (digit crosstalk) is determined by evaluating the integrals of equation 16. No crosstalk occurs between symmetric and skew symmetric functions. Only the following integrals are unequal zero

$$D_{0,cl}=\int_{-1/2}^{1/2} F_0(\theta)\sqrt{2}\cos 2l\pi\theta d\theta$$

$$=\frac{1}{2}\sqrt{2}\,\frac{(-1)^l}{\pi^2 l}\Big[-\mathrm{Ci}\,2\pi(s-l)+\mathrm{Ci}\,2\pi(s+l)-\ln\frac{s+l}{s-l}\Big]$$

$$D_{ck,0}=\int_{-1/2}^{1/2} F_{c,k}(\theta)d\theta$$

$$=\frac{1}{4}\sqrt{2}\,\frac{(-1)^k}{\pi^2 k}\Big[-\mathrm{Ci}\,2\pi(s-2k)+\mathrm{Ci}\,2\pi(s+2k)-\ln\frac{s+2k}{s-2k}\Big]$$

$$D_{ck,cm}=\int_{-1/2}^{1/2} F_{c,k}(\theta)\sqrt{2}\cos 2m\pi\theta d\theta$$

$$=\frac{(-1)^k}{2\pi^2}\,\frac{k}{(k+m)(k-m)}\Big[-\mathrm{Ci}\,2\pi(s-2k)+\mathrm{Ci}\,2\pi(s+2k)-\ln\frac{s+2k}{s-2k}\Big]$$

$$D_{sk,sm}=\int_{-1/2}^{1/2} F_{s,k}(\theta)\sqrt{2}\sin 2m\pi\theta d\theta$$

$$=\frac{(-1)^{k+1}}{2\pi^2}\,\frac{k}{(k+m)(k-m)}\Big[-\mathrm{Ci}\,2\pi(s-2k)+\mathrm{Ci}\,2\pi(s+2k)-\ln\frac{s+2k}{s-2k}\Big]$$

Cix is the cosine integral. Graphs of $D_{0,cl}$ are shown in Fig. 5.

The band limitation causes crosstalk between the digits of adjacent characters. If the character l, represented by the time function $f_l(\theta)$, is transmitted during the time $-1/2\leqslant\theta\leqslant+1/2$, and the character $l+1$ is transmitted during the time $+1/2\leqslant\theta+\leqslant 3/2$, the crosstalk from character l to the digits of character $l+1$ becomes

$$C_0=\int_{-3/2}^{3/2} f_l(\theta)d\theta \tag{22}$$

$$C_{s,k}=\int_{-1/2}^{3/2} f_l(\theta)\sqrt{2}\sin 2k\pi\theta d\theta \tag{23}$$

$$C_{c,k}=\int_{-1/2}^{3/2} f_l(\theta)\sqrt{2}\cos 2k\pi\theta d\theta \tag{24}$$

In the foregoing and in subsequent equations the subscripts c and s stand for cosine and sine, the subscripts k, m denote the respective harmonic. For $f_l(\theta)$ a combination of $F_0(\theta)$, $F_{s,k}(\theta)$, and $F_{c,k}(\theta)$ given by equations 11, 12, 13 must be introduced, depending on which digits are 1 and which are 0. Consider the crosstalk from the constant term of character l to the constant term of character $l+1$.

$$C_{0,0}=\int_{-1/2}^{3/2} F_0(\theta)d\theta=\frac{1}{\pi}\Big[\frac{1}{2\pi s}+2(\mathrm{si}\,4\pi s-\mathrm{si}\,2\pi s)+2\frac{\cos 4\pi s}{4\pi s}-2\frac{\cos 2\pi s}{2\pi s}\Big] \tag{25}$$

Using the asymptotic expansion 18, one may write equation 25 in the following form

$$C_{0,0}\approx\frac{1}{2\pi^2 s}\Big[1+\frac{\sin 2\pi s}{2\pi s}-\frac{1}{2}\,\frac{\sin 4\pi s}{4\pi s}\Big] \tag{26}$$

$C_{0,0}$ is plotted versus s in Fig. 6. For $\alpha=0$ $(s=\alpha+1/2)$, which is the ideal limit of the bandwidth, one obtains a crosstalk of 10%.

The crosstalk between the constant term of character l and the term $\sqrt{2}\cos 2k\pi\theta$ of character $l+1$ is obtained from equations 11 and 24.

$$C_{0,kc}=\int_{1/2}^{3/2} F_0(\theta)\sqrt{2}\cos 2k\pi\theta d\theta$$

$$=\frac{(-1)^k\sqrt{2}}{4\pi^2 k}[\mathrm{Ci}\,4\pi(s+k)-\mathrm{Ci}\,4\pi(s-k)-2\mathrm{Ci}\,2\pi(s-k)] \tag{27}$$

$$\mathrm{Ci}\,x=-\int_x^\infty\frac{\cos y}{y}\,dy\approx 0!\frac{\sin x}{x}-1!\times\frac{\cos x}{x^2}-2!\,\frac{\sin x}{x^3}+3!\,\frac{\cos x}{x^4}\dots \tag{28}$$

Introduction of the asymptotic expansion 28 into 27 yields

$$C_{0,kc}\approx\frac{(-1)^k\sqrt{2}}{4\pi^2 k}\,\frac{\sin 2\pi(s-k)}{2\pi(s-k)}\times[2+\cos 2\pi(s-k)] \tag{29}$$

$C_{0,kc}(-1)^{k+1}k/\sqrt{2}$ is plotted versus $s-k$ in Fig. 7. The crosstalk for $k=1$, $\alpha\geqslant 1/2$ is 1% or less. One may deduce from this result that the crosstalk between adjacent characters is only important for the digits transmitted by the same trigonometric function, constant to constant, $\cos 2k\pi\theta$ to $\cos 2k\pi\theta$, etc.

The crosstalk between the constant in character l and the term $\sqrt{2}\sin 2k\pi\theta$

in character $l+1$ is obtained from equations 11 and 23.

$$C_{0,ks}=\frac{(-1)^k\sqrt{2}}{4\pi^2 k}[4\mathrm{Si}\,2\pi s-2\mathrm{Si}\,4\pi s+\mathrm{Si}\,4\pi(s+k)-2\mathrm{Si}\,2\pi(s+k)+\mathrm{Si}\,4\pi(s-k)-2\mathrm{Si}\,2\pi(s-k)] \quad (30)$$

Using the asymptotic expansion 18 for the sine integral, equation 30 may be expressed in the following form

$$C_{0,ks}\approx\frac{(-1)^k\sqrt{2}}{4\pi^2 k}\times\left[\frac{2\cos 2\pi(s-k)-\frac{1}{2}\cos 4\pi(s-k)}{2\pi(s-k)}-2\,\frac{2\cos 2\pi s-\frac{1}{2}\cos 4\pi s}{2\pi s}+\frac{2\cos 2\pi(s+k)-\frac{1}{2}\cos 4\pi(s+k)}{2\pi(s+k)}\right] \quad (31)$$

The average decline $\sqrt{2}(-1)^k/4\pi^2 k$ is the same as for the cosine terms; the oscillations however are more complicated.

The crosstalk between sine-sine, cosine-cosine, and sine-cosine terms may be computed in the same way, the only difference being that equations 12, 13 contain twice as many terms as equation 11. However, it has been shown by equations 21, 22 that these additional terms yield only small corrections, and equations 26, 29, 31 may thus be used as first approximations for the crosstalk between any orthogonal trigonometric functions of two adjacent characters

$$C_{kc,kc}\doteq C_{ks,ks}\doteq C_{0,0}$$

$$C_{kc,mc}\doteq C_{ks,ms}\doteq C_{0,kc}$$

$$C_{kc,ms}=C_{ks,mc}\doteq C_{0,ks}$$

$$m\neq k \quad (32)$$

The total crosstalk from all digits of character l to the digits of character $l+1$ may now be written from equations 22–24

$$C_0=C_{0,0}+C_{1c,0}+C_{1s,0}+C_{2c,0}+C_{2s,0}+\ldots$$

$$C_{s,k}=C_{0,2k}+C_{1s,sk}+C_{1c,sk}+\ldots C_{sk,sk}+C_{ck,sk}+\ldots$$

$$C_{c,k}=C_{0,ck}+C_{1s,ck}+C_{1c,cb}+\ldots+C_{sk,ck}+C_{ck,ck}+\ldots$$

$$k=1,2,\ldots \quad (33)$$

Since the crosstalk between two different trigonometric functions can be made very small by choosing $s-k=1/2, 1, 3/2, \ldots,$ as may be seen from Fig. 7, one may write as first approximations

$$C_0=C_{0,0}$$

$$C_{s,k}=C_{sk,sk}$$

$$C_{c,k}=C_{ck;ck} \quad (34)$$

The crosstalk from character l to character $l+2$ may be computed in the same way and turns out to be negligible compared with the crosstalk between adjacent characters. This result together with equation 34 justifies the assumption of α being independent of n.

Influence of Noise

Consider a balanced binary character which is represented by digits of value $+1$ and -1 rather than 1 and 0. A signal has the general form

$$f(\theta)=\pm\phi_1(\theta)\pm\phi_2(\theta)\ldots$$

instead of the form in equation 2.

Let the function $\phi_j(\theta)$ be represented by a Fourier series

$$\phi_j(\theta)=\sum_{k=-\infty}^{\infty}a_{jk}\epsilon^{2ik\pi t/\tau}$$

$$a_{jk}=\frac{1}{\tau}\int_{-\tau/3}^{\tau/2}\phi_j(\theta)\epsilon^{-2ik\pi t/\tau}$$

$$E_j=\int_{-\tau/2}^{\tau/2}\phi_j^2(\theta)dt=\tau\sum_{k=-\infty}^{\infty}a_{jk}a_{jk}^* \quad (35)$$

Consider white noise in the frequency domain. Let the frequency band be divided into intervals $\Delta f=f_l-f_{l-1}$. The probability that a sample of noise has a frequency component between f_l and f_{l-1} is independent of l. The amplitudes of the frequency components in any interval Δf have a normal distribution. A sample of noise can be represented by a Fourier series.[1]

$$g(t)=\sum_{l=-\infty}^{\infty}A_l\epsilon^{2il\pi\tau f t} \quad (36)$$

If the function $f(\theta)=\Sigma\pm\phi_j(\theta)$ is transmitted through a noisy channel the signal arriving at the receiver has the form $\Sigma\pm\phi_j(\theta)+g(t)$. The correlation of this signal with the functions $\phi_j(\theta)$ yields E_j or $-E_j$ plus the contribution of $g(t)$.

$$N_{j\tau}=\int_{-\tau/2}^{\tau/2}g(t)\phi_j(\theta)dt=\tau\sum_{k=-\infty}^{\infty}\sum_{l=-\infty}^{\infty}a_{jk}A_l\sin\pi(k+\mu)/\pi(k+\mu)$$

$$\mu=l\Delta f\tau \quad (37)$$

μ becomes a continuous variable for $\Delta f\to 0$ and the sum over l becomes an integral.

For computation of the error probability the distribution of $N_{j\tau}$ normalized by E_j is required. Since the A_l have a normal distribution with the average equal to zero the joint distribution of $N_{j\tau}/E_j$ has the form

$$W(N_{j\tau}/E_j)=(2\pi)^{-1/2}\sigma^{-1}\times\exp[-(N_{j\tau}/E_j)^2/2\sigma^2] \quad (38)$$

A component of $g(t)$ with amplitude A_λ and frequency $\mu=\lambda\Delta f\tau$ correlated with $\phi_j(\theta)$ produces the term $l=\lambda$ in equation 37, the square of this term divided by E_j is the square deviation $\sigma_{A\mu}^2$ from the average $\overline{N}_{j\tau}/E_j=0$.

$$\sigma_{A\mu}^2=\left(\frac{\tau}{E_j}\right)^2A_\lambda A^*\sum_{k=-\infty}^{\infty}a_{jk}\frac{\sin\pi(k+\mu)}{\pi(k+\mu)}\times\sum_{k=-\infty}^{\infty}a_{jk}^*\frac{\sin\pi(k+\mu)}{\pi(k+\mu)} \quad (39)$$

σ^2 is the average of R deviations $\sigma_{A\mu}^2$ for large values of R. The averaging may be performed first for the frequencies then for the amplitudes.

$$\sigma_A^2=\lim_{\Delta f\to 0,R\to\infty}\frac{1}{R}\sum_{r=0}^{R}(\sigma_{A\mu}^2)_r=\int_{-\infty}^{\infty}\sigma_{A\mu}^2d\mu=\left(\frac{\tau}{E_j}\right)^2A_\lambda A_\lambda^*\times\sum_{k=-\infty}^{\infty}a_{jk}a_{jk}^* \quad (40)$$

$$=\frac{\tau}{E_j}A_\lambda A_\mu^*$$

$A_\lambda A_\lambda^*$ is the power dissipated in a unit resistor by the noise components with an

Table I. Bandwidth, Signal Reduction, and Crosstalk Between Characters for a Teletype System

α	W,* Cps	E† 0 Cps	E† 6 Cps	E† 12 Cps	$C_{0,0}$ 0 Cps	$C_{0,0}$ 6 Cps	$C_{0,0}$ 12 Cps
0	15			−1.5%	±4%	±6.5%	±20%
1/2	18			+0.5%	±3.4%	±5%	±10%
1	22			−0.12%	±2.8%	±4%	± 6.5%
3/2	24				±2.2%	±3.4%	± 5%

* All values have to be increased by 50% if start and stop pulses are to be transmitted during one-third of the transmission time.

† Values below 0.1% were not computed.

Table II. Crosstalk Between the Digits of a 5-Digit Character

	1	$\sqrt{2}\sin 2\pi\theta$	$\sqrt{2}\cos 2\pi\theta$	$\sqrt{2}\sin 4\pi\theta$	$\sqrt{2}\cos 4\pi\theta$
1		0	4%	0	8%
$\sqrt{2}\sin 2\pi\theta$	0		0	15%	0
$\sqrt{2}\cos 2\pi\theta$	3.8%	0		0	15%
$\sqrt{2}\sin 4\pi\theta$	0	1.9%	0		0
$\sqrt{2}\cos 4\pi\theta$	4%	0	1.9%	0	

Bandwidth 24 cps ($s=4$), 6 characters per second. The terms above the main diagonal give the crosstalk from the functions on the left to the functions on top; values below the main diagonal give the crosstalk from the functions on top to the functions on the left.

amplitude in the interval $\Delta A_\lambda = |A_\lambda| - |A_{\lambda-1}|$. Hence it follows from equation 40 that the average power of all noise components with amplitudes $|A_\lambda|$ in the frequency interval $-\infty < \mu < +\infty$ equals after integration the average power of all noise components with amplitude $|A_\lambda|$ in the band $-1/2 \leqslant \mu \leqslant +1/2$ before integration. Averaging equation 40 over all amplitudes A_λ yields the average noise power N_n in the band $-1/2 \leqslant \mu \leqslant +1/2$ or $0 \leqslant f \leqslant 1/2\tau$ in real notation

$$\sigma^2 = \frac{\tau}{E_j} N_n = N_n/S_j = N/nS \tag{41}$$

S designates the average power of a signal containing n orthogonal functions of energy E_j each. N is the average noise power in the band $0 \leqslant f \leqslant n/2\tau$, which is the theoretical minimum of the bandwidth required for a transmission rate of n binary digits per time interval τ.

The probability p for $N_j\tau/E_j$ being larger than 1 is obtained from equations 38 and 41

$$p - 1 - \theta(\sqrt{S/2N})$$

$$\theta(x) - \frac{2}{\sqrt{\pi}} \int_0^x e^{-y^2} dy \tag{42}$$

The probability p_n of an error in an n-digit character caused by white, normal noise becomes in first approximation

$$p_n = \frac{1}{2} np \tag{32}$$

The factor 1/2 takes into account the fact that a distortion caused by noise must have the right "direction," that is, only a positive contribution to the digit -1 can cause it to be interpreted as $+1$. The error probability equation 43 is the same as that of a system using amplitude sampling, an ideal low-pass filter of bandwidth $w = 1/2T$, and pulses of the shape $\sin 2\pi wt/2\pi$ wt.[2]

Design Requirements for a Communication System Using Orthogonal Functions

The sine and cosine functions required for the transmitter and the receiver of Fig. 1 may be produced by L-C or R-C oscillators. The phase relationship between sine and cosine may be obtained by integration, use of phase bridges etc. For a Teletype system ($n=5$) transmitting 6 characters per second, the time available to transit one character is 166.6 ms (milliseconds) for synchronous transmission. The five time functions required have thus the following frequencies 0 (constant), 6 cps ($\sin 2\pi t/\tau$, $\cos 2\pi t/\tau$), and 12 cps ($\sin 4\pi t/\tau$, $\cos 4\pi t/\tau$). One has further $T = \tau/n = 33.3$ ms and obtains thus from equation 10 the bandwidth

$$w = (1+\alpha/2.5)/0.066 \text{ cps}$$

The required bandwidths for various values of α are given in Table I together with the reduction E of the received digits (from Fig. 4) and the maximum possible crosstalk $C_{0,0}$ from two adjacent characters (twice the values of Fig. 6). $w = 15$ cps is the theoretical limit of the bandwidth. Table II shows the absolute values of the crosstalk between the digits of a character for a bandwidth of 24 cps.

For a multichannel Teletype system the time functions would be allocated as shown in Table III. Crosstalk and digit reduction for a multichannel system are negligible for all channels except the highest one. The highest channel is also the one that is most affected by synchronization errors. A synchronization error of $\phi = \pm 1/4k$ will change $\sqrt{2} \sin 2k\pi(t/\tau+\phi)$ into $\pm\sqrt{2}\cos 2k\pi t/\tau$ and an output of 0 instead of ± 1 will result. The limit for the tolerable synchronization error in Table III is defined by $E_s = \pm 1/2\phi\tau = \pm\tau/8k$ and is given in milliseconds.

The dependence of the output of the receiver on the accuracy of the synchronization is obtained by evaluating the integrals

$$I_{sk} = 1 \int_{-1/2}^{1/2} \sin 2k\pi\theta \sin 2k\pi(\theta+\phi) d\theta = \cos 2k\pi\phi$$

and

Table III. Frequency Allocation in a Multichannel Teletype System

Channel	Time Function	Frequency, Cps	Limit for Tolerable Synchronization Error*, ± Ms
1	1	0	41.5
	$\sqrt{2}\sin 2\pi t/\tau$	6	20.8
	$\sqrt{2}\cos 2\pi t/\tau$	6	20.8
	$\sqrt{2}\sin 4\pi t/\tau$	12	10.4
	$\sqrt{2}\cos 4\pi t/\tau$	12	10.4
2	$\sqrt{2}\sin 6\pi t/\tau$	18	6.9
	$\sqrt{2}\cos 6\pi t/\tau$	18	6.9
	$\sqrt{2}\sin 8\pi t/\tau$	24	5.2
	$\sqrt{2}\cos 8\pi t/\tau$	24	5.2
	$\sqrt{2}\sin 10\pi t/\tau$	30	4.2
3	$\sqrt{2}\cos 10\pi t/\tau$	30	4.2
	$\sqrt{2}\sin 12\pi t/\tau$	36	3.5
	$\sqrt{2}\cos 12\pi t/\tau$	36	3.5
	$\sqrt{2}\sin 14\pi t/\tau$	42	3.0
	$\sqrt{2}\cos 14\pi t/\tau$	42	3.0

* Synchronization lagging or advancing by the amount shown results in receiving a 0 when a ±1 was transmitted.

Table IV. Encoding, Decoding by Matrix Multiplication

$$\underset{\text{Character Matrix}}{\begin{vmatrix} 0&0&0&0\\ 0&0&0&0\\ 0&0&1&0\\ 0&0&1&1\\ 0&1&0&0\\ 0&1&0&1\\ 0&1&1&0\\ 0&1&1&1\\ 1&0&0&0\\ 1&0&0&1\\ 1&0&1&0\\ 1&0&1&1\\ 1&1&0&0\\ 1&1&0&1\\ 1&1&1&0\\ 1&1&1&1 \end{vmatrix}} \frac{1}{2} \underset{\text{Encoding Matrix}}{\begin{vmatrix} 1&1&1&1\\ 1&1&-1&-1\\ 1&-1&-1&1\\ 1&-1&1&-1 \end{vmatrix}} = \frac{1}{2} \underset{\text{Encoded Character Matrix}}{\begin{vmatrix} 0&0&0&0\\ 1&-1&1&-1\\ 1&-1&-1&1\\ 2&-2&0&0\\ 1&1&-1&-1\\ 2&0&0&-2\\ 2&0&-2&0\\ 3&-1&-1&-1\\ 1&1&1&1\\ 1&0&2&0\\ 2&0&0&2\\ 3&-1&1&1\\ 2&2&0&0\\ 3&1&1&-1\\ 3&1&-1&1\\ 4&0&0&0 \end{vmatrix}}$$

$$\frac{1}{2} \underset{\text{Encoded Character Matrix}}{\begin{vmatrix} 0&0&0&0\\ 1&-1&1&-1\\ 1&-1&-1&1\\ 2&-2&0&0\\ 1&1&-1&-1\\ 2&0&0&-2\\ 2&0&-2&0\\ 3&-1&-1&-1\\ 1&1&1&1\\ 2&0&2&0\\ 2&0&0&2\\ 3&-1&1&1\\ 2&2&0&0\\ 3&1&1&-1\\ 3&1&-1&1\\ 4&0&0&0 \end{vmatrix}} \frac{1}{2} \underset{\text{Decoding Matrix}}{\begin{vmatrix} 1&1&1&1\\ 1&1&-1&-1\\ 1&-1&-1&1\\ 1&-1&1&-1 \end{vmatrix}} = \underset{\text{Decoded Character Matrix}}{\begin{vmatrix} 0&0&0&0\\ 0&0&0&1\\ 0&0&1&0\\ 0&0&1&1\\ 0&1&0&0\\ 0&1&0&1\\ 0&1&1&0\\ 0&1&1&1\\ 1&0&0&0\\ 1&0&0&1\\ 1&0&1&0\\ 1&0&1&1\\ 1&1&0&0\\ 1&1&0&1\\ 1&1&1&0\\ 1&1&1&1 \end{vmatrix}}$$

$$\overset{\text{Encoding Matrix}}{\frac{1}{2}\begin{vmatrix} 1&1&1&1\\ 1&1&-1&-1\\ 1&-1&-1&1\\ 1&-1&1&-1 \end{vmatrix}} \times \overset{\text{Decoding Matrix}}{\frac{1}{2}\begin{vmatrix} 1&1&1&1\\ 1&1&-1&-1\\ 1&-1&-1&1\\ 1&-1&1&-1 \end{vmatrix}} = \overset{\text{Transmission Matrix}}{\begin{vmatrix} 1&0&0&0\\ 0&1&0&0\\ 0&0&1&0\\ 0&0&0&1 \end{vmatrix}}$$

Table V. Encoding, Decoding by Nonsquare Matrices; Transmission Matrix not Unit Matrix

$$\underset{\text{Character Matrix}}{\begin{vmatrix}0&0&0&0\\0&0&0&1\\0&0&1&0\\0&0&1&1\\0&1&0&0\\0&1&0&1\\0&1&1&0\\0&1&1&1\\1&0&0&0\\1&0&0&1\\1&0&1&0\\1&0&1&1\\1&1&0&0\\1&1&0&1\\1&1&1&0\\1&1&1&1\end{vmatrix}}\ \frac{1}{2}\underset{\text{Encoding Matrix}}{\begin{vmatrix}1&1&1\\1&-1&-1\\-1&-1&1\\-1&1&-1\end{vmatrix}}=\frac{1}{2}\underset{\text{Encoded Character Matrix}}{\begin{vmatrix}0&0&0\\-1&1&-1\\-1&-1&1\\-2&0&0\\1&-1&-1\\0&0&-2\\0&-2&0\\-1&-1&-1\\1&1&1\\0&2&0\\0&0&2\\-1&1&1\\2&0&0\\1&1&-1\\1&-1&1\\0&0&0\end{vmatrix}}$$

$$\underset{\text{Character Matrix}}{\begin{vmatrix}0&0&0&0\\0&0&0&1\\0&0&1&0\\0&0&1&1\\0&1&0&0\\0&1&0&1\\0&1&1&0\\0&1&1&1\\1&0&0&0\\1&0&0&1\\1&0&1&0\\1&0&1&1\\1&1&0&0\\1&1&0&1\\1&1&1&0\\1&1&1&1\end{vmatrix}}\ \frac{1}{4}\underset{\text{Transmission Matrix}}{\begin{vmatrix}3&-1&-1&-1\\-1&3&-1&-1\\-1&-1&3&-1\\-1&-1&-1&3\end{vmatrix}}=\frac{1}{4}\underset{\text{Decoded Character Matrix}}{\begin{vmatrix}0&0&0&0\\-1&-1&-1&3\\-1&-1&3&-1\\-2&-2&2&2\\-1&3&-1&-1\\-2&2&-2&2\\-2&2&2&-2\\-3&1&1&1\\3&-1&-1&-1\\2&-2&-2&2\\2&-2&2&-2\\1&-3&1&1\\2&2&-2&-2\\1&1&-3&1\\1&1&1&-3\\0&0&0&0\end{vmatrix}}$$

$$\overset{\text{Encoding Matrix}}{\frac{1}{2}\begin{vmatrix}1&1&1\\1&-1&-1\\-1&-1&1\\-1&1&-1\end{vmatrix}}\times\overset{\text{Decoding Matrix}}{\frac{1}{2}\begin{vmatrix}1&1&-1&-1\\1&-1&-1&1\\1&-1&1&-1\end{vmatrix}}=\overset{\text{Transmission Matrix}}{\frac{1}{4}\begin{vmatrix}3&-1&-1&-1\\-1&3&-1&-1\\-1&-1&3&-1\\-1&-1&-1&3\end{vmatrix}}$$

Table VI. Encoding, Decoding by Nonsquare Matrices; Transmission Matrix a Unit Matrix

$$\underset{\text{Character Matrix}}{\begin{vmatrix}0&0&0&0\\0&0&0&1\\0&0&1&0\\0&0&1&1\\0&1&0&0\\0&1&0&1\\0&1&1&0\\0&1&1&1\\1&0&0&0\\1&0&0&1\\1&0&1&0\\1&0&1&1\\1&1&0&0\\1&1&0&1\\1&1&1&0\\1&1&1&1\end{vmatrix}}\times\underset{\text{Encoding Matrix}}{\frac{1}{\sqrt{8}}\begin{vmatrix}1&1&1&1&1&1&1&1\\1&1&1&1&-1&-1&-1&-1\\1&-1&1&-1&1&-1&1&-1\\1&1&-1&-1&1&1&-1&-1\end{vmatrix}}=\underset{\text{Encoded Character Matrix}}{\frac{1}{\sqrt{8}}\begin{vmatrix}0&0&0&0&0&0&0&0\\1&1&-1&-1&1&1&-1&-1\\1&-1&1&-1&1&-1&1&-1\\2&0&0&-2&2&0&0&-2\\1&1&1&1&-1&-1&-1&-1\\2&2&0&0&0&0&-2&-2\\2&0&2&0&0&-2&0&-2\\3&1&1&-1&1&-1&-1&-3\\1&1&1&1&1&1&1&1\\2&2&0&0&2&2&0&0\\2&0&2&0&2&0&2&0\\3&1&1&-1&3&1&1&-1\\2&2&2&2&0&0&0&0\\3&3&1&1&1&1&-1&-1\\3&1&3&1&1&-1&1&-1\\4&2&2&0&2&0&0&-2\end{vmatrix}}$$

$$\overset{\text{Encoding Matrix}}{\frac{1}{\sqrt{8}}\begin{vmatrix}1&1&1&1&1&1&1&1\\1&1&1&1&-1&-1&-1&-1\\1&-1&1&-1&1&-1&1&-1\\1&1&-1&-1&1&1&-1&-1\end{vmatrix}}\times\overset{\text{Decoding Matrix}}{\frac{1}{\sqrt{8}}\begin{vmatrix}1&1&1&1\\1&1&-1&1\\1&1&1&-1\\1&1&-1&-1\\1&-1&1&1\\1&-1&-1&1\\1&-1&1&-1\\1&-1&-1&-1\end{vmatrix}}=\overset{\text{Transmission Matrix}}{\begin{vmatrix}1&0&0&0\\0&1&0&0\\0&0&1&0\\0&0&0&1\end{vmatrix}}$$

$$I_{ck}=2\int_{-1/2}^{1/2}\cos 2k\pi\theta\cos 2k\pi(\theta+\phi)d\theta$$

$$=\cos 2k\pi\phi$$

For small values of ϕ desired the output is independent of the synchronization error, but the output caused by crosstalk increases linearly with ϕ.

The synchronization of a communication system may be accomplished by transmitting the function $\sqrt{2}\cos 2k\pi\theta$, multiplying it at the receiver with $\sqrt{2}\sin 2k\pi\theta$, and integrating. If receiver and transmitter are in phase the integral over the product will vanish. In the case of an synchronization error of $\pm\phi$ the output of the receiver will be

$$I=2\int_{-1/2}^{1/2}\cos 2k\pi\theta\sin 2k\pi(\theta+\phi)d\theta$$

$$=-\sin 2k\pi\phi$$

The output voltage of the integrator is proportional to the synchronization error ϕ if ϕ is small.

Fig. 1 shows transformers used for adding the various voltages. Resistors may be used as well.

Appendix. Matrix Notation for Encoding and Decoding

It is assumed that the information to be transmitted is represented by binary characters of amplitude 0 or 1. A set of such characters may be regarded as a high matrix. For instance, the 32 characters with 5 digits each used in Teletype transmission form a matrix of 32 rows and 5 columns. For the transmission the 32 characters may be transformed (encoded) into another form, e.g., the functions $f(\theta)$ composed of sine and cosine terms used in this paper. The most general linear transformation that can be performed in Hilbert space is the multiplication by a matrix of finite or infinite rank.

At the receiver the encoded message has to be transformed back into its decoded form. This transformation can again be performed by a matrix multiplication. We have then the "encoding matrix" for encoding the original set of characters and the "decoding matrix" for decoding the encoded characters back into their original form. Since the matrix multiplication is associative, the result of the multiplication of a set of characters by the encoding and then the decoding matrix may also be obtained by multiplying the set of characters by the product of the encoding and decoding matrices. The product of encoding and decoding matrix will be called "transmission matrix."

As an example consider the set of 16 characters with 4 binary digits each. This matrix of 4 columns and 16 rows is multiplied by a square matrix (encoding matrix) with 4 rows and columns; see Table IV. The result is the "encoded character matrix" which shows the sixteen characters in the form in which they are being transmitted. The decoding matrix is so chosen that its product with the encoded character matrix yields the original character matrix. The product of encoding matrix A and decoding matrix B yields the transmission matrix C, which turns out to be the unit matrix, $AB=C=1$. This is the necessary and sufficient condition for the decoded character matrix being equal to the original character matrix.

A case where encoding and decoding matrices are not square is shown in Table V. The encoding matrices of Table V are obtained from the encoding and decoding matrices of Table IV by eliminating the first row. It is noted that the transmission matrix is no longer a unit matrix, but is still a square matrix. Both properties are specializations of the two general rules: 1. The product of two matrices A_{mn} and B_{nm} is a matrix C_{mm}, if m is the number of rows and n is the number of columns. 2. The product of a high matrix and a long matrix cannot be the unit matrix (the product of a long matrix and a high matrix may be the unit matrix).

The practical advantage of using nonsquare matrices is the reduction of digits from 4 to 3 in the encoded character matrix, which means a reduction of time or frequency band for transmission. The price for the reduced bandwidth is evident from the form of the decoded character matrix; the first and the last character are identical and only 31 characters can thus be distinguished. Note that the difference in amplitude of the digits equals 1 for any one of the 30 "good" characters of the decoded character matrix, e.g., 1/4(3+1); 1/4(2+2); 1/4(1+3).

Table VI shows an example of encoding with a long matrix. The encoded characters here have more digits then the original characters. The transmission matrix is a unit matrix and the decoded character matrix is equal to the character matrix.

The transition to matrices of infinite rank can be made in the following way: The product of a matrix A with the terms a_{ik} and a matrix B with the terms b_{jm} yields a matrix C with terms c_{np}. According to the rule of multiplication for finite matrices one has

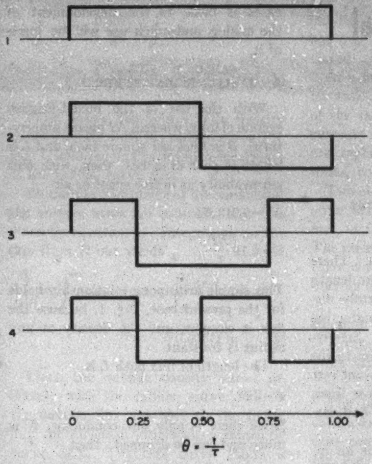


Fig. 8 (left). Set of four orthogonal step functions. Amplitudes are +1/2

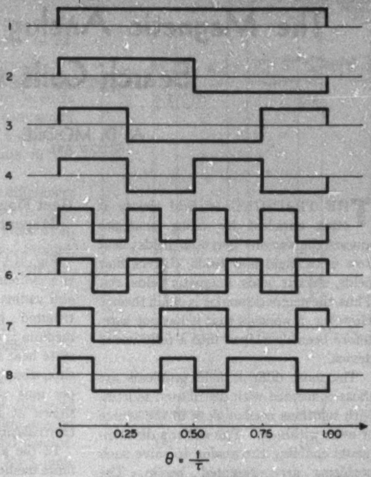


Fig. 9 (right). Set of orthogonal step functions. Amplitudes are $+1/\sqrt{8}$

$$c_{np} = \sum_{i=1}^{s} a_{ni} b_{ip}$$

$$\begin{vmatrix} a_{11} & \cdots\cdots & a_{1s} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ a_{n1} & \cdots\cdots & a_{ns} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ & & \cdot \\ a_{r1} & \cdots\cdots & a_{rs} \end{vmatrix} \begin{vmatrix} b_{11} & \cdots & b_{1p} & \cdot\cdot & b_{1q} \\ \cdot & & \cdot & & \cdot \\ \cdot & & \cdot & & \cdot \\ \cdot & & \cdot & & \cdot \\ \cdot & & \cdot & & \cdot \\ \cdot & & \cdot & & \cdot \\ & & & & \\ b_{s1} & \cdots & b_{sp} & \cdot\cdot & b_{sq} \end{vmatrix}$$

$$= \begin{vmatrix} c_{11} \cdots\cdots & c_{1p} & \cdot\ c_{1q} \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ c_{n1} \cdots\cdots & c_{np} & \cdots\cdots c_{nq} \\ \cdot & \cdot & \cdot \\ & & \\ & & \\ c_{r1} \cdots\cdots & c_{rp} & c_{rq} \end{vmatrix}$$

Let the row $a_{11} \ldots a_{1s}$ become continuously infinite. Instead of the terms a_{1i} one then has continuously variable function $a_1(i)$. Let also the column $b_{11} \ldots\ldots b_{s1}$ become a continuous function $b_1(i)$. The matrix A will then have r rows but continuously infinite columns; matrix B will have q columns and continuously infinite rows; matrix C remains finite and contains terms

$$c_{np} = \int a_n(i) b_p(i) di$$

The integration has to be extended over the whole range of definition of $a_n(i)$ and $b_n(i)$.

As an example, consider the following functions (θ is written rather then i)

$$a_1(\theta) = b_1(\theta) = 1$$

$$a_2(\theta) = b_2(\theta) = \sqrt{2} \sin 2\pi\theta$$

$$a_3(\theta) = b_3(\theta) = \sqrt{2} \cos 2\pi\theta$$

$$a_4(\theta) = b_4(\theta) = \sqrt{2} \sin 4\pi\theta$$

$$a_5(\theta) = b_5(\theta) = \sqrt{2} \cos 4\pi\theta$$

The range of definition is $-1/2 < \theta < +1/2$. Hence it follows

$$c_{np} = \int_{-1/2}^{1/2} a_n(\theta) b_p(\theta) d\theta = 1 \text{ for } n = p$$
$$= 0 \text{ for } n \neq p$$

$$n, p = 1, 2, 3, 4, 5$$

The encoding and decoding matrices used in Table IV may be interpreted as a set of orthogonal step functions as shown in Fig. 8. Leaving out the first quarter of these functions leads to the matrices of Table V. The matrices of Table VI use a subset of the complete set of orthogonal functions shown in Fig. 9.

The encoding and decoding scheme of Fig. 1 has been shown to be described by matrix multiplications. The matrices used are of finite rank if the use time functions are step functions with a finite number of steps and are infinite in all other cases. It follows that the encoding-decoding scheme of Fig. 1 is the most general one that can be devised for linear transformations in Hilbert space.

References

1. Noise and Stochastic Processes (book), **S. O. Rice.** "Mathematical Analysis of Random Noise," Dover Publishing Company, New York, N. Y.

2. The Philosophy of PCM, **B. M. Oliver, J. R. Pierce, C. Shannon.** *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 36, 1948, p. 1324.

3. Communication Theory, **R. M. Fano.** "Communication in the Presence of Additive Gaussian Noise," Academic Press Inc., New York, N. Y.

4. Application of Correlation Analysis to the Detection of Periodic Signals in Noise, **Y. W. Lee, T. R. Cheatham, J. B. Wiesner.** *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 38, 1950, p. 1165.

5. Recent Progress in Applying Information Theory to Digital Transmission Systems, **R. Filipowsky.** *AIEE Transactions*, pt. I (*Communication and Electronics*) vol. 77, 1958 (Jan. 1959 section), pp. 848–55.

The Magnetic Analog with Spiral Search Coils

A. D. MOORE
FELLOW AIEE

THE TEACHING of field theory is often enriched by citing analogies among the various potential fields: heat flow fields, fluid flow fields, current flow fields, electric fields, magnetic fields, etc. Thus, the magnetic analog is old in theory. However, it appears that it has not heretofore been developed into a fully useful device.

The most difficult field problems are those concerned with distributed sources, with solutions needed close to the source or even within it. The writer's developments enabling the analog to solve such problems are presented herein. The analog described applies to 2-dimensional fields in an infinite uniform medium. Some may feel that this seriously limits its scope. Even so, there such problems and they do need solution.

For example, it may not be widely known that, in the gas bubble method of underground gas storage, difficult fluid flow problems arise involving distributed-source phenomena together with an essentially infinite 2-dimensional medium. Large sums are involved, the art of using such storage is widely advanced, and it is rapidly expanding. The magnetic analog might apply to some of these situations.

In any case, policy must always be based on the need for all procurable analogs. There is simply no way of predicting their ultimate utility. A given analog may be useless in one worker's problem in his field; at the same time, it may offer the greatest convenience to the next worker on a different problem in a different kind of work.

The development of the magnetic analog called for the invention of a new form of search coil: the spiral search coil. It may well find applications outside of the analog itself.

Heat Flow Field and Analogous Magnetic Field

Fig. 1(A) presents both cases. There is a cylindrical volume of infinite length and radius R which is a uniformly distributed heat sink. The conducting medium is everywhere uniform. Steady-state heat flows from the medium to the sink, disappearing in it at a constant rate per unit volume. The heat flow lines, shown at the left, are radial, and isothermals are circular.

In the analogous case there is an infinite medium of unit permeability. The cylinder has become a long nonmagnetic conductor-carrying current uniformly distributed. An inward current sets up the magnetic field shown at the right. Flux lines are circular. The magnetic equipotentials outside of the conductor, and the lines-of-no-work inside it are radial. Either field is an inversion of the other. The flux lines of one are equipotentials of the other, and vice versa.

Heat flow has been adopted here for illustrative purposes; it is more familiar to most engineers than some other fields. The particular geometry of the above cases is basic to the development of the analog, and much use will be made of it.

ANALYSIS OF MAGNETIC FIELD

With the use of the mixed-English system of units wherein NI means ampere-turns, B is lines per square inch, and L is length of path in inches; then, with unit permeability as in free space or air

$$NI = 0.313\,BL \tag{1}$$

$$B = 3.19\,\frac{NI}{L} \tag{2}$$

This simple arithmetic relationship holds for the present case, Fig. 1, because the flux is circular, and the density at any radius is constant.

The length of flux path L is

$$L = 2\pi r \tag{3}$$

Since there is only one conductor, N is unity and can be dropped; then

$$B = 3.19\,\frac{I}{2\pi r} = 0.5077 I\,\frac{1}{r} \tag{4}$$

First, analyzing inside the conductor yields

$$\frac{i}{I} = \frac{\pi r^2}{\pi R^2} \tag{5}$$

$$i = I\left(\frac{r}{R}\right)^2 \tag{6}$$

$$B = 0.5077 I\left(\frac{r}{R}\right)^2 \frac{1}{r} = \frac{0.5077}{R^2}\,Ir \tag{7}$$

Next, integrate for the flux ϕ, for unit axial distance, one inch:

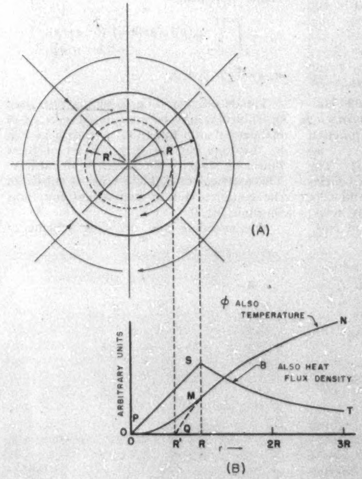


Fig. 1 (left). A—Solid conductor and replacement by smaller thin-wall tube conductor. B—Curve PNM, solid conductor's flux; curve PQMN, tube conductor's flux

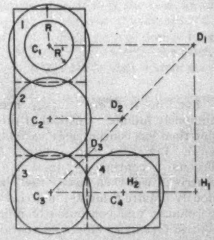


Fig. 2 (above). L-shaped section of distributed heat source; by analogy, section of distributed current

---

Paper 60-129, recommended by the AIEE Computing Devices Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, January 31–February 5, 1960. Manuscript submitted October 26, 1959; made available for printing December 10, 1959.

A. D. MOORE is with the University of Michigan, Ann Arbor, Mich.

The writer is grateful to Profs. L. N. Holland and M. B. Stout, and others of the electrical engineering staff, for assistance on instrumentation, and advice on the use of a power amplifier.

$$\phi = \int_0^r Bdr = \frac{0.5077}{R^2} I \int_0^r rdr$$

$$= \frac{0.5077}{R^2} I \frac{r^2}{2} = \frac{0.2539}{R^2} Ir^2 \quad (8)$$

Thus, inside the source, flux density varies directly with the radius, curve *PS*, Fig. 1; and the flux varies with the square of the radius, curve *PM*.

To analyze for outside of the conductor, the ampere turns are constant, and *B* is given by equation 4. Integrating for the flux from *R* out yields

$$\phi\Big]_R^r = \int_R^r Bdr = 0.5077\, I \int_R^r \frac{dr}{r}$$

$$= 0.5077 I \ln \frac{r}{R} \quad (9)$$

Thus, the outside density varies inversely with the radius, curve *ST*. If the outside flux is added to the inside flux, the variation with radius is shown by curve *MN*. The flux curve *PMN* is continuous; *PM* and *MN* have the same slopes where they meet at *M*.

CORRESPONDING HEAT FLOW FIELD VARIATIONS

It can be shown by similar analyses that the heat field has similar, but inverted, variations. The curve *PST* now displays heat flux density. The curve *PMN* shows the temperature difference between the axis and a point in the field at any radius.

THE ANALOG

It follows then, that if the magnetic case, set up to let the flux out to any radius, be measured, such measurements would stand for, and be in proportion to, the temperature differences between the axis and the same radii in the heat case. There would be no point in doing so for the above case, because it is easily solved analytically. The application appears in due course.

First Replacement: Squares Replaced by Circles

The *L*-shaped polygon of Fig. 2 is made up of four squares. It is the section of a uniformly distributed heat source or sink, infinitely long, and buried in an infinite medium having the same conductivity everywhere. The case could be solved in part by a magnetic analog in which the section becomes that of a solid conductor with uniformly distributed current. The outside field could be explored and mapped, and for some problems this would suffice; for others, requiring solution inside the current, the bus bar type of analog conductor would not do. The inside field cannot be explored.

As a first step toward inside solutions, let each square be replaced by a circle of the same area. The four component currents in the solid square conductors are replaced by equal currents in the round conductors.

Two objections arise. First, differences occur between the new field and the original, in the inside region and close to it. The errors will be treated later. Second, there is an impossibility; the round conductors overlap.

Second Replacement: Solid Round Conductors Replaced by Smaller Tubes

The problem of overlap of the solid conductors is overcome by using smaller tube conductors. Let the solid conductor, Fig. 1, have a radius of $R=1$ inch. The flux measured from the axis out to any radius is shown by the *PMN* curve. It is proposed to replace the solid conductor by a thin-wall tube of radius R' so that, for the same current, the flux from the axis out to any radius is the same as before, for any radius equal to or greater than *R*. One *R*-circle and its R' replacement tube are shown in the square 1, Fig. 2.

To simplify the derivation of R', let the current be such that, at radius *R*, flux density *B* is unity. Also, let the axial length taken be one inch.

The solid conductor's inside flux is found first:

$$B = r$$

$$\phi\Big]_R^r = \int_0^1 Bdr = \int_0^1 rdr = \frac{r^2}{2}\Big]_0^1 = 0.5 \text{ line} \quad (10)$$

The outside flux comes next:

$$B = \frac{1}{r}$$

$$\phi\Big]\frac{r}{R} = \int_1^r Bdr = \int_1^r \frac{dr}{r} = \ln r \text{ lines} \quad (11)$$

The total solid conductor flux beyond *R* is then

$$\phi = 0.5 + \ln r$$

To pass to the tube current, such a current can have no inside flux. The flux around it begins at its surface, at radius R'. It must be required of it that, at unit radius *R*, its flux will be 0.5 line, the same as produced by the solid conductor at that radius. Placing the tube's flux at *R* equal to 0.5 yields

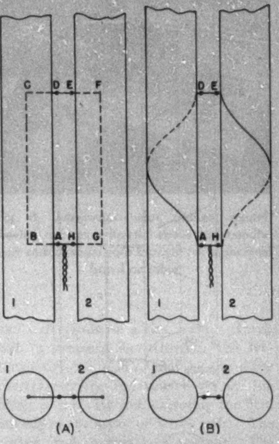


Fig. 3. A—Pair of tube currents with coaxial wire search coils. B—Spiral coils replace the coaxial wires

$$\ln \frac{R}{R'} = \ln \frac{1}{R'} 0.5 \quad (12)$$

$$\ln R' = -0.5$$

$$R' = \epsilon^{-0.5} = 0.606$$

or in general,

$$R' = 0.606R \quad (13)$$

TUBULAR CURRENT'S FLUX

As to the flux linking the axis, Fig. 1, the solid conductor's flux is given by curve *PMN*, whereas that of the tube current is curve *PQMN*. There is perfect agreement along *MN*, for anywhere outside of the solid conductor.

Inside (radii less than *R*) there is disagreement. But this is of no concern whatever, if, as is here done, the measurements are confined to the region outside of *R*.

To sum up, square conductors were replaced by round conductors, which overlapped; when they are replaced by smaller R' tubes, there is ample room for the tubes, and the tube current's field outside of *R* is identical with the round conductor current's field.

COAXIAL-WIRE SEARCH COILS

Fig. 3(A) shows a search coil arrangement to handle two tubes. Tube 1 has a coaxial wire *BC*, brought out radially to terminals *A* and *D* through holes in the tube wall. Tube 2, likewise equipped, has the *EFGH* wire. The wires are joined at *D* and *E*, with twisted leads

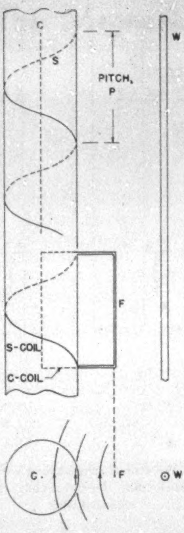


Fig. 4. Spiral search coil. Validity proof uses this figure

brought out from A and H. A rectangular search coil is thus formed, ready to measure flux by fluxmeter if direct currents are used; or by vacuum-tube voltmeter if alternating currents are used. The flux measured would be that which links with the coil.

Suppose tube 2 carries no current. It is interesting to note that the length BC in tube 1 need not then be coaxial: it could be anywhere inside tube 1, and give the same reading. Tube 1 can set up no flux inside itself. However, when two or more tubes are used, coaxial wire locations are necessary, so that net flux between centers may be measured, and ascribed to those locations.

The tube-and-coaxial-wire combination is feasible, but it presents handicaps. First, tubes of finite wall thickness would have to be used; the higher the currents used, the greater the thickness; some error might thus be introduced. In the second place the construction difficulties are formidable.

Third Replacement: Spiral Search Coil and Its Theory

The spiral search coil was invented to evade the construction difficulties mentioned above; in the third replacement, it does away with the coaxial wire coil; see Fig. 3(B). Instead of wires being put inside of tubes, they are simply wrapped around them spirally. The spiral search coil concept has been described to numerous workers familiar with flux measurements; it was new to all of them. It remains to prove that the replacement makes no change in the flux measured.

The theory of the spiral search coil is as follows:

Let the cylinder surface, Fig. 4, be wound uniformly full of spirals to the number of N strands, to make a hollow stranded cable. If each strand carries the current I_S, cable current is NI_S. The cable current sets up a circular flux, some of which links with wire W. Spiral pitch is P. Let ϕ_P be the flux linkage per pitch.

If the stranded cable should be replaced by an equal tube current flowing axially along the cylinder surface, the tube current would cause the same ϕ_p linkage with the wire. This is because axial components of spiral current are identical with elements of the tube current; and peripheral components of spiral current can cause no linkage with wire W.

Next, no change of linkage with W occurs if the NI_S tube current is replaced by an equal current flowing along axis C. Therefore, the axis current gives the same linkage as did the stranded cable current. Since all strands of the cable contribute equally to the linkage, it follows that one strand or spiral S, having current I_S, has the same linkage ϕ_{PS} per pitch with with W, as would be caused by I_S flowing along axis C.

Therefore, for one or more full pitches, the mutual inductance M_{SW} of spiral S with wire W is the same as the mutual inductance M_{CW} of axis wire C with W. Therefore the mutual inductance of wire W with axis wire C is the same as with spiral S. This is to say that when a current in wire is changed, its flux induces equal emf (electromotive forces) per pitch, in S as in C.

In Fig. 4, there are also two closed single-pitch search coils (the outer straight parts, meant to be superimposed, are drawn slightly apart that they may be seen). It is now obvious that current in wire W will cause the same linkage with both these coils, and that either may be used to measure the flux through the C-coil rectangle. This was amply verified by a drastic test made on a special large spiral built for the purpose.

It follows further that the straight parallel length of search wire at F may be replaced at will, by another spiral at any other tube. In Fig. 3, then, the spiral search coil will do the same job as the coaxial-wire coil, and with far greater ease of construction.

Table I. First Replacement Error*

Position	Per-Cent Error
H_1	0.66
H_2	0.67
D_1	0.57
D_2	0.91
D_3	3.5†

* Error in flux measured from center C_3 out to various positions, due to replacement of square by round conductor of equal area; see Fig. 2.
† Approximate.

Spirals must be one or more full pitches in length. Otherwise, it makes no difference whether they are wound to the right or left, or with long or short pitch. All spirals must have the same axial length; lead wires joining two spirals to make a search coil should be held in planes normal to the axis, out to where they can be twisted, so as to have no emf induced in them.

Ordinary search coils often have several, or many, turns. Spiral search coils as described are 1-turn coils, and this form will be retained here. Note, however, that more than one turn can be used. For example, each rod could be wound with two spirals, side by side. The two leads at one end of a rod would be brought out to a terminal board and joined to make a 2-turn search coil; jumpers would be used at each end to achieve the proper in-series connections.

Fourth Replacement: Tubes Replaced by Rods

In the second replacement, a solid round conductor was replaced with a thin tube of smaller radius. The validity de-

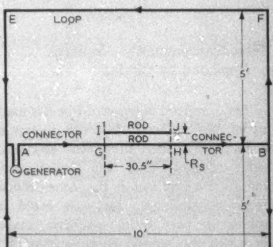


Fig. 5. Layout of one rod element of magnetic analog, with double return loop and one additional rod

Vol
Pt.
196

Table II. Return Loop Errors

	Return Loop Error, Per Cent		
	Single Loop		
Rs, Inches	Roving Coil Outside Loop	Roving Coil Inside Loop	Double Loop
0.26	−0.91	+0.91	
0.52	−0.91	+0.92	
1.05	−1.20	+1.23	
2.10	−1.78	+1.87	+0.040
4.20	−2.74	+3.05	+0.154
8.40	−4.45	+5.30	+0.425

pended in part on the tube's current having no inside flux. If the R' tube is now replaced by a solid rod of the same size, inside flux appears. If coaxial-wire search coils were used, the validity is destroyed. But it was a pleasant surprise to find that if the rod has the spiral coil, validity is restored. The spiral, being wound around circular inside flux, cannot link with it, and ignores it. This advantage was not anticipated when the spiral search coil was conceived.

The rod offers advantages: first, the need for using an extremely thin tube (or else accepting error) is eliminated; second, the rod has superior stiffness and higher current-carrying capacty.

ACCOMMODATING TO THICKNESS OF SPIRAL WIRE

It has been assumed that the spiral wire had no thickness. Real wire has an insulated thickness that is not negligible. Heretofore, rod diameter R' was made to be (0.606 R), but this does not hold for real wire. R' must now be taken as the mean radius of the spiral, with rod radius smaller than R'. If we take R'' as rod radius R'' will be R' minus the radius of the insulated wire section.

The use of the smaller rod entails no error. In fact, any still smaller rod may be used if the spiral is kept coaxial with it. Such a use is considered later.

Summary of Dimension Relationships

The basic dimension in the analog is the size of the original square, which is also the center-to-center distance (as C_1 to C_2); see Fig. 2. Taking this dimension as unity:

Size of square $= 1$
R, radius of circle of same area $= 0.564$
R', mean radius of spiral coil, $0.606 \times 0.564 = 0.366$
R'', rod radius $= R'$ minus wire radius

This order might imply that the design routine is to adopt square size, then fix other dimensions. Instead, rod and spiral wire sizes would be settled upon first, and other dimensions derived therefrom.

Replacement Error

Error appears with the first replacement. The round conductor replaces the square conductor, and their two fields are different, internally and in the nearby region.

Let square 3, in Fig. 2, carry current. It would set up a certain flux between its center C_3 and a position at some radius r. Its round replacement will in general set up a different flux. The writer carried out extensive error studies (not included), the results being given in Table I. It is seen that for nearby centers H_1, H_2, D_1, and D_2, the error is less than 1.0%. Further out, the error is less. The error at the square's corner D_3 is about 3.5%.

Thus the first replacement introduces error for fluxes between centers, but it is very small. These are positive errors; for the positions cited, replacement flux is in excess of flux due to square conductors. The other three replacements, ending in the rod-spiral combination, cause no further errors.

Additional Straight Search Wires

Measurements for additional positions inside the source and at its boundary would sometimes be desirable. These can be had by installing stretched straight search wires at corners of squares, as at D_3 in Fig. 2. Any such wire would act in series with any other, or with a spiral, to form a complete search coil. The flux of square 3 between C_3 and D_3 would be in error by about 3.5%, but this would be the largest error. When wire D_3 is used with rods further away, the error is less.

Layout for One Element of the Analog

A feasible layout for one element (one rod) is shown in Fig. 5. The rod GH is 30.5 inches long, its spiral being wound on for 30 inches. Flux measurements thus are made between the planes GI and HJ. Flexible connectors AG and HB extend the straight part of the current to a distance AB of 10 feet. The current returns by a single loop $BCDA$; or divides equally and returns by double loops $BCDA$ and $BFEA$, whichever design is adopted.

The 1/4-inch rod has an R'' of 0.125 inches. The no. *30* Belden enamelled copper wire for spirals has a diameter of 0.0106 inch. R' is found to be 0.1303 inch.

IJ is another rod, with spiral. The IJ and GH spirals, in series, form a search coil to measure flux through the rectangle $GHJI$. Rod IJ may be one of the current-carrying rods which set up the field; or a roving rod, exploring the field outside of the array of active rods. The roving rod carries no current. The pioneer model is built to the above specifications; see Fig. 6.

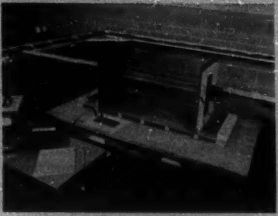


Fig. 6. Magnetic analog pioneer model. Roving rod seen directly above rod bundle. Left ends of spiral coils brought out to terminal board on table

Return Loop Errors

Heretofore, flux computation was simple, the currents being taken as straight and infinitely long. Now, the conductor has the finite length AB, and the current returns nearby. A return loop error will ensue. The error study is much too long to be included, but results are given in Table II. For the case when the roving spiral is in the loop plane, these are the largest errors. For example, when the distance R_s between centers, Fig. 5 becomes 8.4 inches, a single return loop gives errors of −4.45% and +5.30%, depending on whether the roving spiral is outside or inside the loop. The two loops largely cancel their effects when the double loop is used, dropping the error to +0.425%. The double loop also reduces spurious emf due to fields of extraneous currents.

The low error with double return predicts that a single rod would produce a highly circular field out to 8 inches or more. A test on the pioneer model at a 7.5-inch radius, made with eight roving rod positions 45 degrees apart, gave verification: the greatest departure from average flux measured was 0.6%.

Modifying Relative Loop Size. If the rod-and-spiral design as described is retained but return loops are made smaller, loop error will rise. However, a new error study might well show that a double-loop frame size as small as 8 by 8 feet, or even 7 by 7 feet, could give all of the accuracy needed for handling many problems.

Source and Sink Cases. Consider a case in which there is a source, represented by a bundle of rods with current to the right; and a nearby sink of the same total strength, represented by the same number of rods, in a bundle with current to the left. There would be no return loops. All rods would be in series, the connector junctions being made at the terminal block locations.

Principal Design Considerations

The analog inherently has low flux values, measured by 1-turn search coils. These considerations justify the use of long spirals, such as 30 inches. They also indicate the use of alternating current in the rods, with the vacuum-tube voltmeter as the instrument.

The power supply can be 60 cycles if the analog is large enough to carry the currents needed. A large analog makes it easier to get a given degree of construction accuracy.

Rod stiffness is a factor, to get minimum sag and avoid lack of parallelism for horizontal rods supported only at the ends. In a standing-on-end design, with rods vertical, sag would be eliminated.

Bending of rods by electromagnetic forces should be considered. In the illustrative case worked on the pioneer model, and described later, the maximum total lateral force on a rod, at a high momentary current of 100 amperes, was computed to be equal to about one-fifth of the weight of the rod. Spacers can be provided at rod centers, if needed.

Main Features of the Pioneer Model

The pioneer analog was built as indicated in Fig. 5, with double loops laid on a 10- by 10-foot wooden frame. The frame tilts to let the far side rest on the floor. One can walk under the near side. Fig. 6 shows the frame when horizontal.

The rod frame has a base with two head panels rigidly mounted on it. The 1/4-inch Masonite panels, 17 inches wide and 18 inches high, have the smooth sides facing in. The inner faces are $30\frac{5}{8}$ inches apart.

The rods are of 1/4-inch round full-hard brass, $30\frac{1}{2}$ inches long, each with a spiral of no. *30* enamelled copper wire. The spiral was hand-wound by holding the rod at an angle to the oncoming wire and turning the rod. The pitch is 1/2 inch. The 30-inch spiral stops 1/4 inch from each end and is tied down at the ends by wraps of thread. An insulation test at 115 volts eliminated spirals with weak spots in the enamel.

The panels have an array of matched holes in the central areas, and any one rod is positioned and held by a mated pair of holes. The square size (distance between adjacent rod centers) works out to be 0.382 inch. Each panel has 169 holes in a square array of 13×13 holes, spaced 0.382 inch on centers.

Each rod is extended one inch at each end by 1/4-inch brass studs which go through the panel holes. The inner end of a stud screws into the end of the rod. The stud's outer end is drilled to receive the connector wire, which is soldered into it. A connector such as *HB*, Fig. 5, is braided no. *12* copper, rubber-covered. The outer end terminates in a spade lug, for attachment to terminal blocks mounted at *A* and *B*. The double loops are pairs of no. *16* braided copper, rubber-covered, the pairs ending in spade lugs permanently attached to the terminal blocks.

To mount a rod, it is held to hole positions. A loose stud-connector is picked up, the stud is slipped through its hole, and is screwed into the rod end; this is repeated for the other end. Connections are then made to the terminal blocks.

This model is designed for 60 cycles. Normal currents of 10 to 20 amperes are easily carried. When unusually low fluxes are measured, currents can go to 50-ampere levels (or more) for long enough to get readings.

Screw connections were adopted to make sure of handling such currents. Use of higher frequency and lower current levels would permit use of a slip fit of stud into rod, and banana-plug endings for the connectors. This would make it much faster, in changing from one setup to another.

The inner panel faces have horizontal and vertical lines scratched on the smooth surfaces to form a grid with 0.382-inch spacings. The grid locates the roving spiral when exploring the outside field. The holes in an array are centered on intersections.

When the roving spiral comes within the hole array, it is located by special studs going through holes; a stud extension fits smoothly with a hole drilled in the end of the dummy rod which carries the spiral.

A spiral and its leads are all one piece of wire. When rods are being inserted during a setup, their longer leads at one end are brought out and connected to a small terminal board that duplicates the setup. The shorter leads at the other end are all connected together. The pioneer model is large, but the design permits disassembly. The loop frame stacks against the wall when not in use.

Voltage Equation

The voltage equation for a magnetic analog is given below. It applies when one rod carries current, and its spiral, together with a roving spiral, forms a search coil. With the use of a sine wave:

I = amperes (effective)
f = frequency
L_S = axial length of spirals, inches
R' = mean radius of spiral on rod, inches
r = radius to center of roving spiral, inches
E = volts induced

$$E=3.19IfL_S(10)^{-8}\ln\frac{r}{R'} \qquad (14)$$

In the pioneer model, let one rod be in use, with the roving rod one hole away from it. R' is 0.1303, and r is 0.382. With 60 cycles, and the high momentary current of 100 amperes

$$E=3.19\times100\times60\times30(10)^{-8}\ln\frac{0.382}{0.1303}=0.00618 \text{ volt}$$

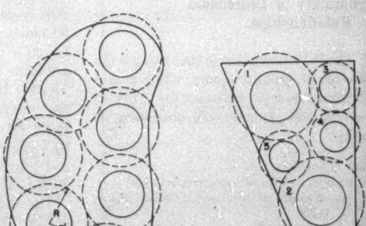


Fig. 7. A—Irregular source section replaced by equal rod elements. B—Another section, replaced by two sizes of rod elements

Adaptation of Analog to Irregular Sections

When the distributed source section divides into squares with regular spacings, the construction described suffices. The handling of iregular sections is indicated in Fig. 7. In Fig. 7(A), the seven R-circles, whose totalized area equals that of the original section, could be drawn to parcel out the section very well indeed. This amounts to the first replacement. Each circle now replaces a polygon rather than a square. Next, the R'-circles (rod-and-spiral combinations) replace the R-circles. This is a case when the section may well be carved into equal sized elements.

Other cases occur, Fig. 7(B), where a limited number of elements gives a better fit by using two, or more, sizes. Here, the larger areas 1 and 2 are equal, and the smaller areas 3 through 6 are all equal. With uniform distribution in the source, rod currents are made proportional to the areas they represent.

If rods (or tubes) of the larger size are available, they can be used, with the spirals wound directly on them. Or smaller rods all of the same size can be used throughout: a rod having to carry a large spiral can be built up by slipping a thin-wall brass tube concentrically over it, the tube to carry the spiral. Commercial tube and wire sizes can be found that will combine to bring R' very close to a required value.

If additional measurements are made with straight search wires, they should be placed where R-circles do not overlap. The 3.5% replacement error found for D_3 at the square's corner, Fig. 2, will now, with polygons instead of squares, be greater for some positions, and less for others.

Construction for Irregular Spacings. The regular hole arrays described could not locate rods with irregular spacings. Flexibility of hole positions is needed. The central areas of both panels would be cut out, making, say, square openings were rod bundles are located. Masonite hole-pieces would be cut to fit the openings as inserts. The inserts would be drilled for the studs at the positions required.

Nonuniform Distributions

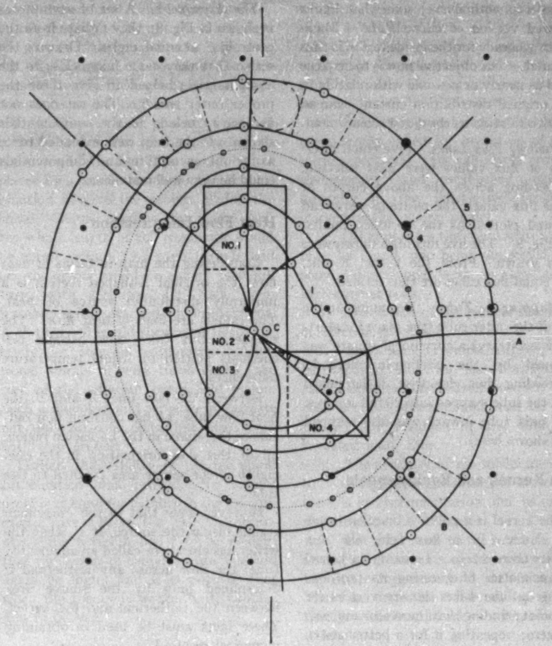


Fig. 8. Computed field map for L-section case; analog should produce same field. Analog flux dashed line in tube 3-4 agrees with computed field

Potential field problems occur in which the source has a nonuniform distribution. For example, heat production with the section, Fig. 7(A), might not be uniform. The rate of heat production assignable to each polygon would be known, and individual rod currents made proportionate to these values.

Numerical Solution of a Case

Before building the pioneer model, the case of Fig. 2 was worked out on paper, to see how much information the proposed analog could yield and how useful it would be in solving the field. The original section, Fig. 8, was given the coarsest breakdown possible, being treated as only four square current components. It was assumed the replacements had been made, with the rod-spiral combinations in use. The total flux per axial inch of no. 1, for example, from center out to any radius was then given by only one equation, equation 9, with R in the denominator changed to R'.

The steps leading to the computed map, Fig. 8, will be outlined. They have much to do with the use of the analog itself. Also, the map will be used to indicate in a brief way some routines involved in putting the map to use, to obtain desired answers from it.

No. 2 was chosen as the reference rod, it being foreseen that the current centroid C and the kernal K would fall within it. The positions taken were also kept on the coarse, limited basis, being confined to the 42 black dots; see Fig. 8. The routine was to compute the flux from no. 2 center out to any position, due to the 4-rod currents.

Two families of curves (not included) were then plotted. An H-family of seven curves came from flux-versus-horizontal position, going across the horizontal rows. A V-family of six curves were of flux-versus-vertical position, along the vertical rows. Curves from diagonal plots can be had for additional points, and one such was used.

These curves are much alike, broadly speaking, being upswept either way from where they go through lows within the source. It is most important to give the correct trend to the low part of such a curve. Here, the rod-spiral treatment is not used in sweeping in the low part, for this would give a sort of skewed version of the broken curve $PQMN$. Instead, for this particular purpose, the first replacement is reverted to, and the sweep is in a smooth curve, somewhat like a skewed version of curve PMN. Maximum accuracy for the inside region is thus secured. The objective now is to map the field as nearly as possible with relation to the original distribution current, instead of with relation to the 4-rod replacement.

Plotting Flux Lines. Five equal intervals of flux values were next selected. Proceeding across the above curves at these flux values, the positions were read off and plotted as the encircled points; see Fig. 8. The five flux lines shown were then drawn. From the kernel K out, five equal-flux tubes are thus formed.

Mapping the Tubes. Beginning at the top of the outer tube (the place for starting is arbitrary) a curvilinear square was mapped by the circling-in method.[1] Proceeding thus clockwise, it happened that the tube mapped into 26 full squares. The next tube inward was also mapped (not shown here).

The Kernel, and Equipotentials

The kernel is a point of inaction inside the source. In a flux field, the flux density there is zero. Locating the kernel was a matter of guessing its position; laying out the 4-flux density vectors for the point; finding that the vector sum was not zero; repeating it for a better guess, etc. Actually, the trial-and-error work quickly ended in a very good location at K.

The eight lines radiating from the kernel came next. Inside the source they are lines of no work; outside, they are equipotentials. The lines were to be at equal potential differences. If 3.25 is obtained by dividing the 26 squares of the outer tube by 8, it follows that these lines should cross the tube at intervals of 3.25 curvilinear squares. An arbitrary beginning is made somewhere, and the lines thus located. Since graphical work is subject to some error, the next tube inward, also mapped, was likewise treated, and both tubes were allowed to compromise in deciding the locations.

The lines were then prolonged inward to K, with two criteria in mind: first, that the map must be orthogonal; second, that the lines should divide the source area into eight equal areas. Some trial-and-error work was required to do this sufficiently well.

Be it specifically noted that the lines do not arrive at K with equal angles. The belief is often met that they should do so; instead, the several angles can be fantastically unequal, as this case illustrates.

The Asymptotes. A set of asymptotes is shown in Fig. 8; they radiate from the centroid C, at equal angles. Drawing the set so that they came fairly close to the equipotentials helped to give these the proper outer trends. The set does not necessarily belong to the equipotentials shown. If the map were extended twice as far out, asymptotes and equipotentials could be very well matched.

Heat Flow Interpretation

Considering the map in terms of heat flow, the original L-shaped section is a uniformly distributed source of heat. The radials are lines of heat flow. The magnetic flux lines now become isothermals plotted for equal temperature differences.

Between the flow lines A and B, the heat flow tube AB has constant flow outside of the source in the Laplacian region, all of this flow originating in the one-eighth of the source area devoted to this tube.

However, note that in the Poissonian region inside the source, it is what the writer has elsewhere called an augmented tube; the flow across any isothermal is determined only by the source area between the isothermal and the kernel. These facts must be used in obtaining numerical results.[2]

Use of Heat Flow Map. Let k be the medium's heat conductivity, in these terms: If one degree centigrade difference exists between the faces of a 1-inch cube, then k watts of (parallel) heat flow occurs through the cube. Passing to the map, consider any curvilinear square volume of 1-inch axial depth (into the paper) and faced by any curvilinear square. The conductance of this volume is the same as that of the 1-inch cube.[1] Curvilinear square volumes can be combined in series, or parallel, or both, in terms of Ohm's law. Thus, if temperature differences are known, heat flows can be computed, or vice versa.

The inner part of tube AB is partly mapped into squares. One of these squares is subdivided into four curvilinear squares. The temperature drop across this square, for example, can be taken as due to the heat originating in the area between the subdivider isothermal, and the kernel.[2]

Case Solved on Analog

The case of Fig. 2, already solved on paper on a coarse basis and giving the map shown in Fig. 8, was set up on the analog, Fig. 6. However, the analog was set up in terms of the next finer degree of subdivision: 16 rods were used, or four for each of the original four squares. The rod bundle occupied the central area of the hole array. The sixteen rods were in series. The roving rod is shown directly above the rod bundle.

The current was 50 amperes, used momentarily, at 60 cycles. The highest search coil voltages were obtained when the roving rod went to the corners of the hole array, these then being around 60 to 80 millivolts.

Plots of the data in terms of H- and V-families gave highly consistent curves. One of the experimental flux lines is shown dashed about midway of the tube defined by computed lines 3 and 4. The analog-produced line is highly consistent with the computer lines. Other parts of the two maps likewise agree. The experimental points are enclosed in the smaller circles close to or on the dashed line.

Use of Higher Frequencies

Rod current would vary inversely with frequency, for the same search coil voltage output. Therefore, speaking of the pioneer model, a high momentary current of 50 amperes at 60 cycles could be replaced by a steady current of 5 amperes, 600 cycles. Since the model can carry 20 amperes, this current at 600 cycles would give four times as many millivolts to measure on the vacuum-tube voltmeter.

Readings were taken of the case described above, at 10 amperes, 680 cycles, by the use of a rotating generator as the source. A very good check was obtained with readings from the 60-cycle measurements.

Another test was made at 2 amperes, 4,600 cycles, with a power amplifier as the power supply. The check was poor, the fluxes within the rod bundle close to the kernal being below expectation. This suggests that proximity effect was causing a redistribution of rod currents within the rod sections. Approximate calculations of proximity effect bore this out. It is concluded that the design of the pioneer model permits frequencies up to about 700 cycles, without need to worry about proximity effect.

Using hard brass for the rods not only gives stiffness; it also gives desirably high resistivity, helpful in minimizing proximity effect at higher frequencies. At 680 cycles, the ohms reactance for the analog as set up was found to be 4.9 ohms. This converts to 0.43 ohm at 60 cycles.

Vol.
Pt.
196

Types of Use for Analog

At least two kinds of uses for the analog can be envisaged, here to be called types 1 and 2. Type 1, the "map" type, is illustrated by Fig. 8 and the discussion of it. The problem requires a complete, or rather complete, analysis. The operation of the analog leads to a flux map. Any convenient rod current value can be used, as long as the map is obtained. Once the map exists, its magnetic flux origin is forgotten, and it becomes a thermal (or other) map, standing on its own merits, and ready to be used in terms of heat data put into it in order to get answers from it, i.e., in type 1 uses, the analog current chosen need have no relation whatever to the ultimate use of the map.

In type 2, the "direct" type, the analog would be set up and used on a specific case, to give immediate answers. For example, for a given rate of heat production, it is required to know the temperatures at a few particular locations. Appropriate conversions would then be made, so that when analog current is fixed at the right value to represent heat produced, emf measured could at once be converted to degrees of difference.

In the foregoing, the word "current" applies when all rods have the same current. The word becomes plural, when a case requires rod currents to be individually set.

Other Uses of Spiral Search Coil

The spiral search coil may find applications outside of the analog. For example, large current-limiting reactor coils continue to be studied and improved. They are 3-dimensional and axisymmetric. Forces tending to disrupt them are high and may go higher as generator capacities rise. It is important to know the flux densities within the reactor, so that force computations can be made.

A small-scale model of a proposed coil could be built. The conductor radius, to scale, would be R. This would be replaced by a smaller conductor-spiral combination to give the proper R' for the spiral. Before winding the model coil, the whole length of the conductor would be spiral wound. As each turn is laid on, the spiral would be cut at each full turn, with terminals for each turn-length of spiral attached and brought out. A 48-turn coil would then have 48 spirals, ready for internal exploration. Additional spirals or wire turns could take care of the adjacent outer field region. Such a model should serve to investigate not only air-core coils, but also those with iron present.

Earlier, the validity of the rod-spiral combination was proved. No similar proof is at hand for validity in the axisymmetric case. From the practical standpoint, the writer holds it to be evident that the center-to-center accuracy would again be high; and it is possible that when the theory is worked out, validity may be proved.

Conclusions

1. The magnetic analog has been changed from a theoretical concept to a workable reality. Two-dimensional potential fields of heat flow, fluid flow, current flow, etc., in an infinite uniform medium, can be solved by setting up the analogous magnetic field.
2. The analog described achieves solutions inside of distributed sources, as well as outside of them.
3. Uniformly distributed sources of regular or irregular boundaries can be handled, and for nonuniform distributions.
4. The development of a usable magnetic analog called for invention of the spiral search coil. Its validity is proved theoretically, and established experimentally in the illustrative problem worked on the pioneer model.
5. Other uses for the spiral search coil may be found, one being suggested for some 3-dimensional work on very large, high-current coils.
6. The design of the pioneer model is completely described.
7. The analog is simple, rugged, can be kept to moderate size limits, and can be operated on 60-cycle power. The cost is low. Measurements can be suspended and resumed at any time, or repeated at any time.
8. The accuracy of the model appears to be so high that the principal source of error is in instrumentation, due to using the vacuum-tube voltmeter.
9. For watt-hour meter research, the analog's rod currents could replace driving flux components and their images.[3] The flux map from the analog would then give the map of eddy currents in the disk.

References

1. MAPPING TECHNIQUES APPLIED TO FLUID MAPPER PATTERNS, A. D. Moore. *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 71, Jan. 1952, pp. 1–6.

2. SOAP FILM AND SANDBED MAPPER TECHNIQUES, A. D. Moore. *Journal of Applied Mechanics*, New York, N. Y., vol. 17, Sept. 1950, pp. 291–98.

3. EDDY CURRENTS IN DISKS: DRIVING AND DAMPING FORCES AND TORQUES, A. D. Moore. *AIEE Transactions*, vol. 66, 1947, pp. 1–11.

◆

Discussion

Walter E. Rogers (University of Washington, Seattle, Wash.): I should like to submit an alternative mathematical proof of Professor Moore's somewhat surprising conclusions regarding the spiral search coils. I shall show that the net flux contributed to the coil by the spiral *surrounding* a current of circular symmetry is independent of the way in which the spiral is wound. On the other hand, I shall show that a spiral *with a uniform pitch*, wound on a supporting cylinder external to the currents producing the field, is effectively located on the axis of that cylinder.

Referring to Fig. 4, let us compute the flux through search coil S assuming that the cylinder which supports the spiral carries a current which is distributed with circular symmetry. While the common procedure of evaluating the flux through any convenient surface bounded by the search coil produces the same result, the use of Stokes' theorem and the equivalent line integral of the vector potential is simpler and more illuminating.

As is customary, let us define the vector potential **A** by

$$\mathbf{B}=\text{curl }\mathbf{A} \tag{15}$$

Reference to equation 7 will demonstrate that a suitable vector potential for points outside the current-carrying region has only an axial component and is given by

$$A_z=c\log r \tag{16}$$

where r is the distance from the axis of the current and c is a constant depending on the total current, its direction on the axis and on the system of units.

The flux through the search coil is given by the line integral of A around the perimeter of the search coil.

$$\phi=\oint \mathbf{A}\cdot d\mathbf{S} \tag{17}$$

with

$$d\mathbf{S}=\mathbf{i}\,dx+\mathbf{j}\,dy+\mathbf{k}\,dz \tag{18}$$

The first segment of the integral is path AD in Fig. 3(B). Since this segment is at constant distance from the axis, the vector potential is a constant. Where R is the radius of the supporting cylinder

$$A_z=c\log R \tag{19}$$

and the differential is exact.

$$\int_A^D \mathbf{A}\cdot d\mathbf{S}=\int_A^D A_z dz=c\log R(D-A) \tag{20}$$

This contribution to the flux is therefore independent of the shape of the spiral, and depends only on its axial length. Indeed, this leg of the search coil need only lie on the surface.

Referring again to Figs. 3(B) and 4, the horizontally drawn segments of the search coil are perpendicular to the vector potential and contribute nothing to the line integral. The remaining portion is the straight segment F of Fig. 4. Here A_z is also a constant, since r is a constant, and

$$\int_D^A \mathbf{A}\cdot d\mathbf{S} = c \log r(A-D) \tag{21}$$

Combining equations 20 and 21 into equation 17, yields

$$\phi = c(A-D)\log \frac{r}{R} \tag{22}$$

which is equivalent to equation 9.

We turn now to the case in which the spiral segment is external to the sources of the field. This is illustrated by Fig. 4 if the current sources are confined to wire W. To draw a conclusion, we need only to compute the line integral along the spiral path of the S coil. Let r again be the radial distance to the axis of the currents, now in wire W, and let g be the vector distance between axes. Let the parametric equations of the spiral be

$$|\mathbf{r}-\mathbf{g}| = \rho \tag{23}$$

where ρ is the radius of the supporting cylinder, and

$$\theta = \frac{2\pi}{P} z \tag{24}$$

We find here that the line integral

$$\int_A^D A_z dz = \int_A^D c \log r dz \tag{25}$$

depends on the path, since r is a variable. Using the law of cosines and the path of equation 24, we transform the variable to obtain

$$\int_A^D A_z dz = \int_0^{2\pi} \frac{cP \log g}{2\pi} d\theta + \int_0^{2\pi} \frac{cP}{2\pi} \log\left[1+\frac{\rho^2}{g^2}-\frac{2\rho}{g}\cos\theta\right] d\theta \tag{26}$$

The second definite integral on the right is zero when the bracketed quantity is always positive, as it is in this case. Therefore

$$\int_A^D A_z dz = cP \log g \tag{27}$$

We interpret this result in the following way: The contribution of the uniform spiral segment is the same as though it were a straight wire of length equal to its pitch and located at the center of its supporting cylinder. This interpretation arises because in the result, $c \log g$ is the vector potential on the axis.

While a result equivalent to the spiral could be achieved by a straight segment lying on the surface of the cylinder at distance g from one wire W, the circular symmetry will be destroyed if other wires are present.

Circular symmetry products results which would not be expected from intuition alone. Another example is the fact that the mutual inductance of circuits of long, parallel round wires depends only on their spacing and not on their size.[1] Neither does it depend on the distribution of the current, as long as circular symmetry is preserved.

Reference

1. Introduction to Electric Fields (book), W. E. Rogers. McGraw-Hill Book Company, Inc., New York, N. Y., 1954, pp. 314–17.

A. D. Moore: As Prof. Rogers says of the spiral search coil, circular symmetry produces results which would not be expected from intuition alone. This is confirmed by the way it came into existence. The constructional difficulties presented by the coaxial search wire arrangement, Fig. 3(A), were so distasteful that the writer was driven to grasp at any other idea that might show promise. This may explain why the spiral search coil has come along so belatedly, in spite of the wide use of spiral wire forms in many examples of electrical engineering.

One is grateful to Prof. Rogers for his analyses, and especially for the alternative proof of validity of the spiral search coil. He lends emphasis to the fact that the search coil spiral should be uniform. However, this should not lead to the conclusion that extreme accuracy is required. In the magnetic analog, for example, due care should insure that hand-wound spirals will suffice, if several to many pitches are used. One may repeat the hope that the spiral search coil will find uses other than that covered in the paper.

Shall an Electron Tube or a Semiconductor Device Be Used?

E. E. SCHENEMAN
MEMBER AIEE

S. K. WALDORF
MEMBER AIEE

A WIDE variety of electron tubes is now available to satisfy the needs of almost every application, but new types are still appearing as the field of electronics continues to enlarge. Also available are many kinds of semiconductor devices, although they are not as numerous as the types of electron tubes. With such a large range of operating characteristics available in both classes of electron devices, a designer of electronic equipment often is obliged to consider both types and make a choice of either an electron tube or a semiconductor device. The proper selection may not be immediately apparent, although each kind of electron device has a number of distinctive characteristics which can serve as aids in arriving at a reasonable decision.

It is the objective of this paper to present in convenient form a summary and comparison of numerical data in order of magnitude, not in exact figures, on both electron tubes and semiconductor devices which will serve as a general guide for those areas where both types of devices might be considered as being worthy of consideration.

General Considerations

In any one application, the designer necessarily must make the choice between an electron tube and a semiconductor device after making his own evaluation of relative costs, marketability, fabrication and maintenance conditions, and the factors enumerated in the accompanying tables. It is important to note that semiconductor devices have the following distinctive characteristics:

1. Require no cathode heater power.
2. Certain types perform well at very low voltage.
3. Generally they have low mass, as little as 0.1 gm (gram) for diodes and 0.3 gm for transistors, and small volume, 0.4 cc (cubic centimeters) minimum, with consequent excellent capability of resisting vibration and mechanical shock.

On the other hand, electron tubes also have several distinctive characteristics:

1. Very high electric resistance between tube elements is available for isolation of circuits.

Paper 60-74, recommended by the AIEE Electronics Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., January 31–February 5, 1960. Manuscript submitted March 18, 1959; made available for printing December 8, 1959.

E. E. Scheneman is with the Westinghouse Electric Corporation, Baltimore, Md. S. K. Waldorf (deceased) was with the Westinghouse Electric Corporation, Baltimore, Md.

The authors wish to acknowledge the information supplied by the Baltimore Divisions, the Electronic Tube Division, and the Semiconductor Department of the Westinghouse Electric Corporation.

Table I. Satisfactory Environmental Conditions for Semiconductor Devices and Electron Tubes

			Electron Tubes		
	Semiconductor Devices*		Thermionic		
Environmental Condition	Germanium	Silicon	Vacuum	Gas Filled	Cold Cathode
Junction or envelope temperature	−65 C to 85 C	−65 C to 175 C	−65 C to 500 C (Ceramic)	−55 C to 90 C	−55 C to 90 C
Altitude	No limit		Sea level to 80,000 ft; heating and flashover may reduce rating		
Relative humidity	0 to 100%		0 to 100%		
Maximum vibration	30 g at 20 to 2,000 cycles per second†				
Maximum mechanical shock	40 g at 10 milliseconds; 500 g at 1 millisecond				
Method of mounting	Usually soldered; some plug-in		Soldered or plug-in		
Thermal shock	0 C to 100 C		0 C to 100 C		
Failure method	Sudden or gradual		Sudden or gradual		
Nuclear radiation	See reference 1		See reference 2		

* See Figs. 1 and 2.

† See explanation of tables in text.

2. Operating temperature can exceed about 175 C (degrees centigrade), the upper limit for most semiconductor devices.

3. Very high voltages can be rectified, converted, or utilized.

Electron tubes also are now made in small sizes with volumes as low as 0.58 cc and masses of 1.6 grams.

There are often considerable differences in the criteria of acceptability used by designers of equipment for use by the Department of Defense, by industry, or by the general public. Equipment used by the first of these groups usually has the most stringent conditions to satisfy, with some reduction of requirements for industrial equipment and least severe conditions for consumer products. Correspondingly, tubes and semiconductor devices are made with varying degrees of reliability and sturdiness and at different levels of cost. Thus, each designer must rely upon his own judgment and experience, but it is hoped that the tables presented here will serve as check lists to expedite the preliminaries of design procedure.

Explanation of Tables

The data given in the accompanying tables and graphs were compiled from manufacturers' data, published literature, and specifications of the United States Department of Defense. The numerical data in the tables are based on characteristics of units generally available commercially at the end of 1958. It must be kept in mind that the limiting values of operating conditions are being extended almost continually by new developments in both electron tubes and semiconductor devices.

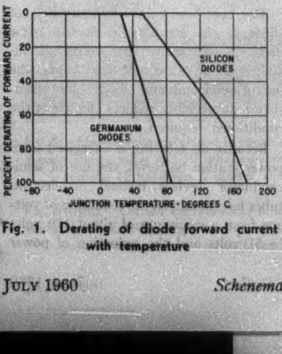


Fig. 1. Derating of diode forward current with temperature

It is believed that the information in the tables in general is self-explanatory; however, a few comments follow for clarification.

Table I. Environmental

Temperature. Operating temperature has a marked effect upon the performance of semiconductor devices. For example, although the nominal upper temperature limit for silicon units is 175 C, at this temperature both the current and power ratings drop to zero, as shown in Figs. 1 and 2. Thus, for practical purposes, the permissible maximum temperatures for semiconductor devices are appreciably under the nominal values given in Table I. However, certain units are now being manufactured where the zero rating point is at about 200 C.

Generally, it is not recommended that electron tubes with soft glass envelopes or seals be operated above about 175 C for maximum life and reliability. Some tubes with ceramic insulation are rated for 500 C operation, with certain developmental types rated for 750 C.

Altitude. The voltages applied to, and the heat dissipation in, most semiconductor devices usually are so low that there are no altitude limitations on the operation of these devices.

However, the appreciable voltage and heat dissipation associated with many electron tubes places limitations on the altitude or atmospheric pressure at which they can operate. Reduced ambient gas density may require reduced voltages to prevent external flashovers and reduced rating and losses to limit tube temperature to a practical level. The cooling effect of the ambient gas is reduced with reduction of gas density at increasing altitudes.

Relative Humidity. Humidity may reduce insulation resistance and affect both types of electron devices in much the same way, although the greater heat dissipation in electron tubes is sometimes helpful in drying insulation. Semiconductor devices must be hermetically sealed to prevent rapid deterioration.

Maximum Vibration and Mechanical Shock. Although Table I shows the same limits for both types of electron devices, semiconductor devices are inherently better able to withstand physical stresses, and manufacturers and users report that semiconductor devices in general are appreciably more resistant to vibration and mechanical shock than are electron tubes. The values in the table are those often specified for tests, although some transistors are now being vibration tested at 100 *g* (gravitational acceleration).

Method of Mounting. The high temperature of molten solder may change the characteristics of, or even damage, semiconductor devices. When mounting them in circuits by soldering, it is advisable to provide temporarily some form of heat sink on the lead being soldered, between the device and the soldered joint. This usually will limit the temperatures in the device to tolerable values.

Thermal Shock. In R-C (resistance-capacitance) type oscillators using transistors, thermal shock has been noted to have a marked effect on the frequency stability. It has been found that a change in temperature of about 20 C over a period of a few minutes causes the fre-

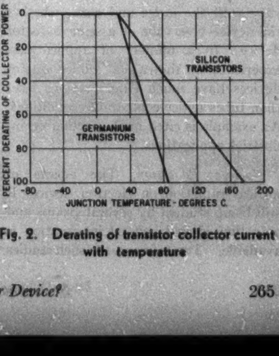


Fig. 2. Derating of transistor collector current with temperature

Table II. Satisfactory Electrical Operating Conditions for Hot-Cathode Vacuum Tubes and Transistors

Operating Condition	Transistors		Hot-Cathode Vacuum Tubes
	Germanium	Silicon	
Input resistance	10 to 5×10^5	10 to 5×10^5	10^5 to 10^9
Input signal power	10^{-8} to 1	10^{-8} to 1	10^{-13} to 10^3
Input signal current	10^{-5} to 1	10^{-5} to 1	10^{-7} to 2
Input signal voltage	10^{-6} to 5	10^{-6} to 5	10^{-6} to 10^3
Output or load impedance	1 to 5×10^5	1 to 5×10^5	1 to 5×10^6
Maximum output signal power			
Class B amplification	50	85	10^5
Switching	10^3	2×10^3	
Maximum output signal current	12	10	6
Maximum output signal voltage	10^2	2×10^2	10^5
Maximum power gain, db*	40	40	70
Maximum voltage gain, db			45
Maximum current gain, db	10^2	10^2	
Cathode heater power	none	none	5×10^{-2} to 3×10^3
Plate or collector voltage	10^{-1} to 2×10^2	10^{-1} to 2×10^2	12 to 10^5
Plate or collector current	10^{-6} to 12	10^{-6} to 10	10^{-6} to 1
Amplifier operating frequency	0 to 5×10^8	0 to 10^8	0 to 10^{10}
Maximum oscillator operating frequency	8×10^8	10^8	10^{10}
Ability to withstand surges	poor	poor	good
Maximum variation between units of same type	400% in current gain	400% in current gain	30% in transconductance
Temperature coefficients	current gain may change	current gain may change	stable after warm up

* Decibels.

Table III. Satisfactory Electrical Operating Conditions for Tubes and Semiconductor Devices Used as Switches, Regulators, or Rectifiers

Operating Condition	Semiconductors		Tubes	
	Germanium	Silicon	Vacuum	Gas or Vapor
Maximum peak inverse voltage	4×10^2	2×10^3	1.5×10^5	2×10^4
Inverse current	10^{-6} to 10^{-2}	10^{-5} to 10^{-2}	Zero	Zero.
Maximum d-c output voltage	2×10^2	1.5×10^2 single 1.5×10^3 stack	7.5×10^4	10^4
Maximum average output current	500	300	1	700
Maximum power efficiency	99.9%	99.9%	98%	99+%
Junction or envelope temperature	*	*	−65 C to 200 C	−55 C to 90 C
Maximum operating frequency as detector	Point contact 10^9 Junction 2×10^6	Point contact 10^{10} Junction 60×10^6	10^8	10^3
Regulated voltage		1 to 6×10^2		60 to 2×10^4

* See Figs. 1 and 2.

quency to drift excessively, whereas the same temperature variation over a few hours has only a minute effect. In electron tubes this effect is not present, probably because of the heat liberated by the cathode heaters.

Failure Method. Both types of electron devices may fail suddenly or gradually. The life and reliability of a device vary with the usage and class, whether the device be a tube or a semiconductor device. In a number of applications, experience has shown that semiconductor devices have much longer service lives than tubes under comparable conditions. An example is cited in the section on industrial applications.

Nuclear Radiation. The effects of nuclear radiation on electron devices is still being studied by several groups and only limited and general information is available. Two references to such studies are given.[1,2] In general, semiconductors suffer deleterious effects at lower levels of radiation than do tubes.

TABLE II. ELECTRICAL: VACUUM TUBES AND TRANSISTORS

Input Resistance. The input resistance rather than the input impedance is listed in the table because the former is relatively fixed for any one electron device, whereas the latter varies with the operating frequency

Voltage Gain. Does not apply to transistors which are essentially current amplifiers.

Current Gain. Does not apply to vacuum tubes which are essentially voltage amplifiers.

Amplifier Operating Frequency. The zero lower limit in the table signifies that direct currents and voltages can be amplified.

Ability to Withstand Surges. Semiconductor devices are quite susceptible to surge voltages and are not suitable in circuits where component failures, or maintenance or operating procedures can introduce high-voltage transients. A recent experimental study indicates that the ability of transistors to withstand surges can be improved by some changes in construction details.

Table III compares the characteristics of tubes and semiconductor devices used as switches, regulators, or rectifiers.

Examples of Choices for Equipment

CONSUMER PRODUCTS

The use of transistor hearing aids has become prevalent because of the remarkably small size and light weight of these devices. The prolonged battery life makes this type of hearing aid additionally attractive because users do not feel compelled to carry a spare battery. The long life and resistance to mechanical shock of transistors also has greatly improved reliability. For these reasons, transistors have the outstanding advantage in this field.

In television receivers, the electron tube predominates because of the high frequencies involved, the closer tolerances maintained in the manufacture of tubes, the high voltages required for the picture tube, and the higher cost of semiconductor devices for the relatively large number of units needed.

INDUSTRIAL APPLICATIONS

A typical industrial application in which semiconductors are used throughout is in a carrier-current transmitter-receiver for electric power system relaying. The use of semiconductors and miniature components reduced the size so that it can be mounted in a standard relay case instead of requiring a 72-inch-high cabinet, as can be seen in Fig. 3. This permitted it to be mounted on the same switchboard with the associated relays, which facilitates over-all relay testing.

The heat losses are less than 10% of those in the comparable electron tube unit. Where many such units are installed in an air-conditioned room, the use of semiconductor diodes and transistors appreciably reduces the size of the conditioner required.

Because all relaying is energized from power station batteries, the use of semiconductor devices instead of electron tubes has permitted the reduction of voltage from a minimum of about 129 volts to 51 volts and the reduction of power

Vol.
Pt. 1
1960

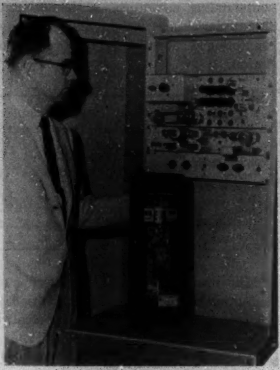


Fig. 3 (left). Transistorized carrier transmitter-receiver in black case is approximately 1/20th the size of the equivalent tube assembly including cabinet

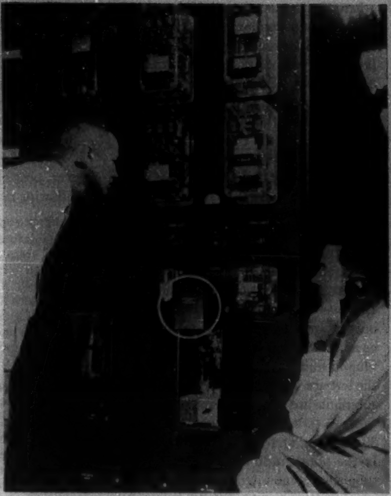


Fig. 4 (right). Unit circled is a voice adapter unit for voice communication over the relaying carrier channel. It is small and sturdy enough to be carried in a technician's tool kit

requirements by 90%. This has permitted the use of appreciably smaller batteries and battery chargers.

Semiconductor devices have made possible an entirely new feature in this carrier-current equipment: a small, compact 3-transistor voice adapter unit for use with the relaying transmitter-receiver. This unit, shown circled in the center of Fig. 4, permits voice communication over the relaying channel for installation, test, and maintenance purposes when it is plugged into an octal socket on the relaying unit. Its small size, 6 by 4 by $3^1/_2$ inches, and the sturdiness of the transistors make it practicable for a technician to include the unit in his kit of tools and have it available for use at any location. A feature of the equipment is the priority given to the relaying functions, which automatically and instantaneously take precedence over voice communication when necessary. That is, if a conversation is in progress when a relay operation occurs, the speech signal is blocked and full transmitter output produced for relaying purposes. The low-impedance input of transistors has made it possible to block the modulator with a small change in base current. This greatly simplifies circuit requirements. The low impedance of transistors also has made it possible to make a single transistor perform two functions in the transmitter-receiver. The output transistor is used as a modulator when transmitting and as an audio amplifier when receiving. Relay contacts effectively switch the transistor from transmitting to receiving, and vice versa, utilizing the low-impedance characteristic of the transistor.

The anticipated increased reliability of the semiconductor devices as compared to electron tubes has been realized in field installations of the relaying equipment, resulting in less maintenance. On one power system the transistor equipment has operated more than 3,200,000 transistor hours with only one failure, in contrast to average tube life of 15,000 hours, which would have required 130 tubes for the same length of time. These figures take into account that there are less tubes required for the equivalent tube equipment.

Although the foregoing example is an outstanding one for which semiconductor devices are particularly appropriate, there are numerous other industrial applications for which electron tubes are especially suited, such as microwave equipment

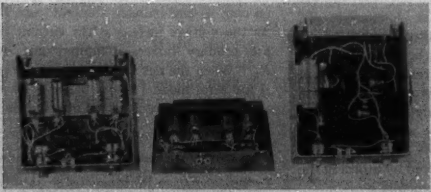


Fig. 5. Plug-in units on left and in center are the same, each containing 4 subminiature tubes and 10 other components. Plug-in unit on right contains 12 transistors, 2 diodes, and 33 other components

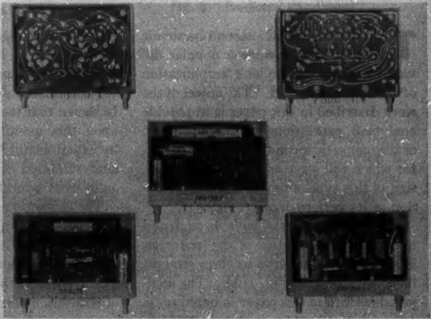


Fig. 6 (right). On left are two views of a d-c amplifier; on right, two views of a stabilizing amplifier; in center, a transistorized monostable multivibrator where the frequencies are higher than the limits for semiconductor devices; welders, where the currents to be controlled are large and the time intervals are short; and induction heating, where the power requirements are large and high voltages are needed for the output tubes.

MILITARY APPLICATIONS

To attain the design objectives in military aircraft equipment of minimum size and weight with capability for withstanding vibration and mechanical shock, subminiature electron tubes and semiconductors are used, each type of device selected for its distinctive electrical characteristics. In general, electron tubes are used as output amplifiers, semiconductor devices as amplifiers and switches. In Fig. 5, the plug-in unit in the center and the one on the left are identical, each containing four subminiature tubes, the tubes chosen because output signal voltages greater than 30 volts and up to 200 volts were required. In the photograph can be seen the metal tube shields mounted on a metal support which in turn is welded to the frame of the plug-in unit. The frame of the unit fits snugly in its mounting in the assembled equipment, all of these intimate metal-to-metal contacts being necessary to dissipate the approximately 5 watts of cathode heater power in the four tubes. It will be noticed that no components other than tubes are mounted on the metal heat-conducting strip, or "heat sink."

The transistor unit shown on the right in Fig. 5 is $4^1/_2$ by $5^1/_8$ by 7/8 inches, only about 15% larger in volume than the electron tube unit shown in the photograph, yet it contains several times as many circuit components as are in the tube unit. The airborne equipment housing these two types of plug-in units is not mounted on shock isolators.

Semiconductor devices were used wherever possible in the equipment for the following reasons:

1. The typical power supply voltage for subminiature tubes is 150 volts as compared to 30 volts for transistors. Assuming the same current drain, the power consumption is reduced 80% by the use of transistors, in addition to the elimination of cathode heater power of approximately $1^1/_4$ watts per tube. Thus, the heat to be dissipated by the equipment is minimized when transistors are used.
2. The small mass and volume of semiconductor devices.
3. Transistors permit superior timing circuit accuracy when they are used as switches because their collector saturation potential, on the order of 100 millivolts or less, yields greater uniformity in the performance of transistors of the same type. However, allowances must be made in the design for the effect of temperature changes on the performance of semiconductors.

Fig. 6 shows three computer plug-in units for military equipment which is not airborne, yet is made as compact as possible by the use of silicon semiconductor devices. The two amplifiers shown in the photograph, together with a chopper unit not shown, replace a tube amplifier. The outstanding advantages of the new units are savings in size, weight, and power consumption as shown in the following tabulation:

	Tube	Semiconductor
Weight, gm	359	197
Volume, cubic inches	14.3	13.8
Power, watts		
B+	2.38	1.46
B−	3.25	1.47
Cathode heaters	8.45	None
Totals	14.08	2.93

The marked reduction in power consumption has permitted the elimination of blowers for cooling, and the use of smaller circuit components.

The new unit has proved more reliable, will operate over a wider temperature range, has better zero and drift characteristics, and can be manufactured with better and more uniform quality than the former tube units. The use of printed circuits has contributed to better quality control by eliminating some wiring errors during assembly.

References

1. EFFECT OF NUCLEAR RADIATION ON SELECTED SEMICONDUCTOR DEVICES, G. L. Keister, H. V. Stewart. *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 45, July 1957, pp. 931–37.
2. SURVEY ON THE EFFECTS OF NUCLEAR RADIATION ON ELECTRON TUBE MATERIALS. Stevens Institute of Technology, Hoboken, N. J., U. S. Department of the Army contract DA 36-039 SC 73146.

An Analysis of the Operation of the Magnetic Second-Harmonic Modulator

B. W. JALBERT
ASSOCIATE MEMBER AIEE

THE MAGNETIC second-harmonic modulator has become a popular device for the detection and amplification of low-level d-c signals. The object of the work described in this paper is to provide analytical expressions of the operation of the magnetic second-harmonic modulator which will serve as a basis for the design of this type of device and the prediction of its performance.

Previous analyses[1,2] have been based upon the representation of the hysteresis loop of the core material by three straight lines, as shown in Fig. 1(A). The analysis described in this paper is based upon the rectangular hysteresis loop as shown in Fig. 1(B). It is believed that this assumption is more realistic in view of the core materials now available. It will be shown that the analytical results based upon this assumption are corroborated by experimental results. The method of analysis is that employed by Storm in the analysis of saturable reactors.[3]

The magnetic second-harmonic modulator circuit upon which this analysis is based is shown in Fig. 2. It is assumed that the cores have a hysteresis loop as in Fig. 1(B), that they have the same value of saturation flux density, and that the sinusoidal supply voltage is greater than twice the voltage required to saturate one core.

To understand the operation, first assume that the supply voltage is at its maximum negative value. Core A will be at negative saturation while core B is at positive saturation. A positive control current will decrease the net ampere turns on core A, and increase those on core B. As the voltage increases, the current in the primary increases, limited only by R_p. When the net ampere turns on core A are zero, mode 1 starts. The voltage across core A is transformed into the load

---

Paper 60-143, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., January 31–February 5, 1960. Manuscript submitted August 17, 1959; made available for printing December 8, 1959.

B. W. JALBERT is with the General Electric Company, Auburn, N. Y.

The author wishes to express his gratitude to J. B. McFerran and F. J. Ellert for their assistance in the preparation of this analysis.

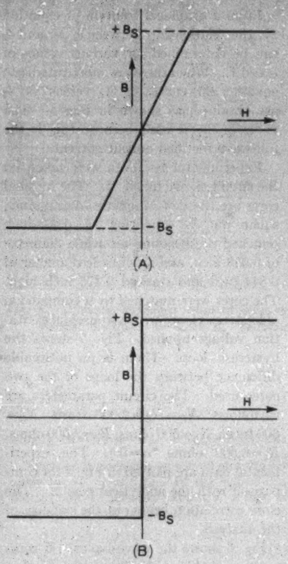


Fig. 1. Straight-line hysteresis loop

A—With finite slope
B—Rectangular

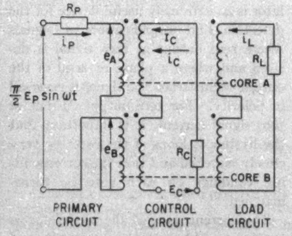


Fig. 2. Circuit diagram

and control windings. When the net ampere turns on core B become zero, that core comes out of saturation and mode 2 is initiated. During this mode, the load current is constant and proportional to control current. When core A reaches saturation, the voltage across that core becomes zero and mode 3 commences. During this mode, the voltage across core B is transformed into the load. Since this voltage is opposite in polarity to that across core A, the load current reverses. Mode 4 starts when core B gets into negative saturation. In mode 4, the load current is zero. The waveshapes of voltage, current and flux are shown in Fig. 3. As the supply voltage decreases, core B comes out of saturation first. Since the voltage is going negatively and since the polarity of core B is opposite to that of core A, the voltage across the load winding of core B will be the same as that across the load winding of core A in the previous half-cycle. Thus, the output waveshape repeats itself each half-cycle. Since it contains both positive and negative portions, the output contains only even harmonics, the largest of these being the second harmonic. A reversal in control current will cause a reversal in phase of the load current.

The voltage across core A during interval I is

$$e_A=\frac{1}{1+\alpha}\left(\frac{\pi}{2}\,E_p\sin\,\omega t+\frac{N_C}{N_P}I_CR_P\right) \tag{8}$$

where α is the ratio of the primary circuit resistance to the parallel combination of reflected control circuit resistance and load resistance. This may be expressed as

$$\alpha=\frac{R_P}{N_P{}^2}\left(\frac{N_L{}^2}{R_L}+\frac{N_C{}^2}{R_C}\right) \tag{5}$$

Interval 1 commences when positive volt-seconds are applied to core A to effect a flux change in that core. Interval 1 then starts when the voltage across core A, equation 8, is zero and increasing positively. This occurs at the angle θ_1, which is determined from

$$\sin\,\theta_1=-\frac{N_CI_C}{N_P\frac{\pi}{2}E_P/R_P} \tag{9}$$

The voltage across the primary of core B during interval 2 is

$$e_B=-\frac{1}{2}\left(\frac{\pi}{2}E_P\sin\,\omega t-\frac{1}{\alpha}\frac{N_C}{N_P}I_CR_P\right) \tag{17}$$

Likewise, this interval starts when the quantity is zero and increasing negatively.

The phase angle at which this occurs is θ_2, and is determined from

$$\sin\,\theta_2=\frac{1}{\alpha}\,\frac{N_CI_C}{N_P\frac{\pi}{2}\frac{E_P}{R_P}} \tag{18}$$

The phase angle at which core A goes into positive saturation is θ_3; that at which core B goes into negative saturation is θ_4. The determination of these angles will be discussed later.

The voltage across core A during interval 2 is

$$e_A=\frac{1}{2}\left(\frac{\pi}{2}\,E_P\sin\,\omega t+\frac{1}{\alpha}\,\frac{N_C}{N_P}I_CR_P\right) \tag{16}$$

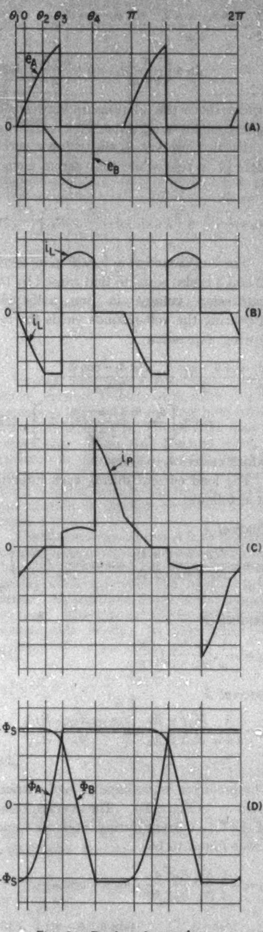


Fig. 3. Predicted waveshapes

A—Gate voltages
B—Load current
C—Primary current
D—Core fluxes

The flux excursion in core A from negative saturation to positive saturation is caused by the volt-seconds applied to that core in intervals 1 and 2. This excursion is equal to that which occurs when the saturation voltage, $\pi/2E_s\sin\omega t$, is applied for one half-cycle. Equating these volt-seconds and introducing the excitation factor $\bar{x}$ yields the following relationship:

$$\frac{2}{\bar{x}}=\frac{\cos\theta_1-\cos\theta_2}{1+\alpha}+\frac{\cos\theta_2-\cos\theta_3}{2}+$$
$$\sin\theta_2\left(\frac{a}{1+\alpha}(\theta_2-\theta_1)+\frac{\theta_3-\theta_2}{2}\right) \quad (27)$$

which shows the relationship between the phase angles θ_1, θ_2, and θ_3.

During interval 3, the voltage across core *A* is zero while that across core *B* is

$$e_B=-\frac{1}{1+\alpha}\left(\frac{\pi}{2}E_P\sin\omega t-\frac{N_C}{N_P}I_CR_P\right) \quad (25)$$

The flux excursion in core *B* in intervals 2 and 3 is also equal to that caused by the saturation voltage in one half-cycle. Equating the volt-seconds yields the following expression:

$$\frac{2}{\bar{x}}=\frac{\cos\theta_2-\cos\theta_3}{2}+\frac{\cos\theta_3-\cos\theta_4}{1+\alpha}-$$
$$\sin\theta_2\left[\frac{\theta_3-\theta_2}{2}+\frac{\alpha}{1+\alpha}(\theta_4-\theta_3)\right] \quad (28)$$

which relates θ_4 to θ_2 and θ_3.

The load current during each interval is as follows:

Interval 1

$$i_L=-\frac{N_L}{N_P}\frac{1}{1+\alpha}\left(\frac{\pi}{2}\frac{E_P}{R_L}\sin\omega t+\frac{N_C}{N_P}I_C\frac{R_P}{R_L}\right) \quad (7)$$

Interval 2

$$i_L=-\frac{1}{\alpha}\frac{N_C}{N_P}\frac{N_L}{N_P}I_C\frac{R_P}{R_L} \quad (15)$$

Interval 3

$$i_L=\frac{1}{1+\alpha}\frac{N_L}{N_P}\left(\frac{\pi}{2}\frac{E_P}{R_L}\sin\omega t-\frac{N_C}{N_P}I_C\frac{R_P}{R_L}\right) \quad (24)$$

The predicted waveshape of load current is shown in Fig. 3(B). The average value of the load current during intervals 1 and 2 was found to be

$$I_L=-\frac{1}{4}\frac{N_L}{N_P}\frac{E_P}{R_L}\left[\frac{4}{\bar{x}}-(\cos\theta_2-\cos\theta_3)+\right.$$
$$\left.(\sin\theta_2)(\theta_3-\theta_2)\right] \quad (30)$$

The average load current during interval 3 was found to be

$$I_L=\frac{1}{4}\frac{N_L}{N_P}\frac{E_P}{R_L}\left[\frac{4}{\bar{x}}-(\cos\theta_2-\cos\theta_3)+\right.$$
$$\left.(\sin\theta_2)(\theta_3-\theta_2)\right] \quad (32)$$

These equations show that the total average load current each half-cycle is zero; hence no d-c component of current appears in the load circuit. Since the positive and negative portions of load current appear in each half-cycle of the supply frequency, the output contains only even harmonics.

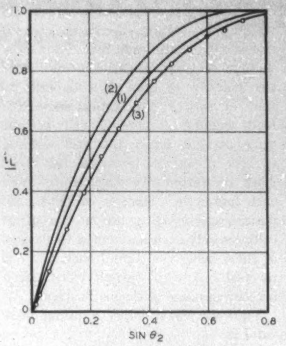


Fig. 4. Per-unit load currents as function of sin θ_1 with $\alpha=0.1$

Curve 1—$\bar{x}=4$
Curve 2—$\bar{x}=6$
Curve 3—Experimental data at $\bar{x}=4$

When the signal current is such that $\theta_2=\theta_3$ (and for larger values of signal current), core *A* will experience its total flux change before core *B* comes out of saturation. For this condition, the average voltage which is transformed into the load circuit for the flux excursion of each core is the saturation voltage. The maximum average value of each positive and negative portion of load current is

$$I_{L,max}=\frac{N_L}{N_P}\frac{E_S}{R_L} \quad (37)$$

That the derived equations may be expressed on a dimensionless basis, the following per-unit quantities are defined.

Per-unit load current (ratio of actual load current to maximum value of load current):

$$\bar{i}_L=\frac{I_L}{\dfrac{N_L}{N_P}\dfrac{E_S}{R_L}} \quad (33)$$

Per-unit control ampere turns (ratio of control ampere turns to maximum load ampere turns):

$$\bar{a}=\frac{N_CI_C}{N_L\dfrac{N_L}{N_P}\dfrac{E_S}{R_L}} \quad (34)$$

From equation 32, the per-unit load current is

$$\bar{i}_L=+\frac{\bar{x}}{4}\left[\frac{4}{\bar{x}}-(\cos\theta_2-\cos\theta_3)+\right.$$
$$\left.(\theta_3-\theta_2)\sin\theta_2\right] \quad (36)$$

From equations 18 and 5, the per-unit control ampere turns is

$$\bar{a}=\frac{\pi}{2}\bar{x}\left(1+\frac{N_C^2}{N_L^2}\frac{R_L}{R_C}\right)\sin\theta_2 \quad (35)$$

From a graphical solution of equation 27, the relationship between θ_1, θ_2, and θ_3 can be determined for various values of α and $\bar{x}$. When these are substituted into equation 36, curves of $2\bar{i}_L$ versus $\sin\theta_2$ are obtained, as shown in Fig. 4. The value of $2\bar{i}_L$ is used because this is the full-wave rectified output current.

Experimental test data were taken on the circuit shown in Fig. 2. The toroidal cores are those of Magnetic Metals, Inc., square-mu *80* consisting of 0.004-inch punched rings having an inside diameter of 0.633 inch, and an outside diameter of 0.844 inch and stacked 0.375 inch high. The cores were matched by a comparison of hysteresis loops with a sinusoidal excitation voltage applied. Fig. 5 shows the hysteresis loop. There is no noticeable difference between the loops of the two cores used. The circuit parameters are as follows: $N_P=N_L=2{,}000$ turns, $N_C=600$ turns, $R_P=500$ ohms, $R_c=2{,}000$ ohms, $R_1=6{,}500$ ohms, $\alpha=0.1$. The experimental data are plotted in Fig. 4 for comparison with the analytical results. The close correlation bears out the validity of the analysis.

Fig. 6 shows the waveshapes. A comparison of this with Fig. 3 shows that the waveshape is as predicted.

Conclusions

The magnetic second-harmonic modulator is an extremely useful device for the detection and amplification of signals where zero output at zero signal is desired and where a phase reversal of the output occurs when the signal is reversed in polarity. For zero output current at zero signal current, it is important that the hysteresis loops of the cores be very closely matched and that the corresponding windings on each core be turns-matched.

The current gain of the modulator is nonlinear and lower in value than that of a series saturable reactor. However, for low values of signal current, the gain is nearly linear.

Appendix

The circuit under analysis is shown in Fig. 2. Consider core *A* originally at negative saturation and core *B* at positive saturation. The first interval of operation occurs from the time core *A* comes out of saturation until core *B* comes out of saturation.

The net mmf (magnetomotive force) on core *A* during this time is zero. The circuit equations during this period are

$$e_B=0 \quad (1)$$

$$\frac{\pi}{2}E_p\sin\omega t=e_A+i_PR_P \quad (2)$$

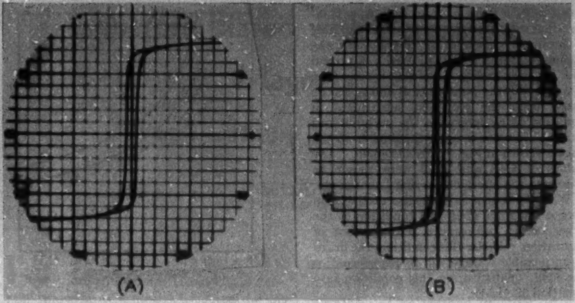


Fig. 5. Dynamic hysteresis loops of cores used. A—Core A: calibration 510 gausses per vertical division. B—Core B: calibration 0.007 oersted per horizontal division

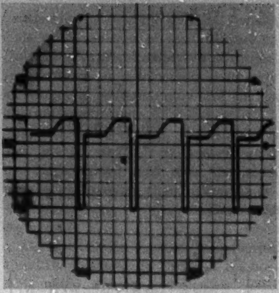


Fig. 6. Load current with excitation factor of $\bar{x}=4$

$$e_A=-\frac{N_P}{N_L}i_LR_L=-\frac{N_P}{N_C}i_CR_C \tag{3}$$

$$\Sigma mmf_A=0=N_Pi_P+N_CI_C+N_Ci_C+N_Li_L \tag{4}$$

These equations can be solved simultaneously for i_p, i_L and e_A. In the solution, a dimensionless quantity appears, which is designated α and is defined as

$$\alpha=\frac{R_P}{N_P{}^2}\left(\frac{N_L{}^2}{R_L}+\frac{N_C{}^2}{R_C}\right) \tag{5}$$

This quantity is the ratio of primary resistance to the parallel combination of reflected control circuit resistance and reflected load resistance. Utilizing equation 5, the simultaneous solution of equations 1 through 4 yields the following:

$$i_P=\frac{1}{1+\alpha}\left(\frac{\pi}{2}\frac{E_P}{R_P}\sin\omega t-\frac{N_C}{N_P}I_C\right) \tag{6}$$

$$i_L=-\frac{N_L}{N_P}\frac{1}{1+\alpha}\left(\frac{\pi}{2}\frac{E_P}{R_L}\sin\omega t+\frac{N_C}{N_P}I_C\frac{R_P}{R_L}\right) \tag{7}$$

$$e_A=\frac{1}{1+\alpha}\left(\frac{\pi}{2}E_p\sin\omega t+\frac{N_C}{N_P}I_CR_P\right) \tag{8}$$

Interval 1 commences when positive volt-seconds are applied to core A. It can be seen that this occurs when equation 8 is equal to zero. This is at the angle θ_1 which is determined from

$$\sin\theta_1=-\frac{N_CI_C}{N_P\frac{\pi}{2}\frac{E_P}{R_P}} \tag{9}$$

Interval 2 occurs from the time core B comes out of saturation until core A goes into saturation.

During this period, the circuit equations are

$$\frac{\pi}{2}E_p\sin\omega t=e_A-e_B+i_PR_P \tag{10}$$

$$e_A+e_B=-\frac{N_P}{N_L}i_LR_L=-\frac{N_P}{N_C}i_CR_C \tag{11}$$

$$\Sigma mmf_A=0=N_Pi_P+N_CI_C+N_Ci_C+N_Li_L \tag{12}$$

$$\Sigma mmf_B=0=-N_Pi_P+N_CI_C+N_Ci_C+N_Li_L \tag{13}$$

From these equations the following are derived:

$$i_P=0 \tag{14}$$

$$i_L=-\frac{1}{\alpha}\frac{N_C}{N_P}\frac{N_L}{N_P}I_C\frac{R_P}{R_L} \tag{15}$$

$$e_A=\frac{1}{2}\left(\frac{\pi}{2}E_p\sin\omega t+\frac{1}{\alpha}\frac{N_C}{N_P}I_CR_P\right) \tag{16}$$

$$e_B=-\frac{1}{2}\left(\frac{\pi}{2}E_P\sin\omega t-\frac{1}{\alpha}\frac{N_C}{N_P}I_CR_P\right) \tag{17}$$

Interval 2 commences when negative volt-seconds are first applied to core B, i.e., when equation 17 is equal to zero. The angle at which this interval starts is θ_2 and is determined from

$$\sin\theta_2=\frac{1}{\alpha}\frac{N_CI_C}{N_P\frac{\pi}{2}\frac{E_P}{R_P}} \tag{18}$$

During interval 3, which is the time from which core A goes into saturation until core B goes into saturation, the circuit equations are

$$e_A=0 \tag{19}$$

$$\frac{\pi}{2}E_P\sin\omega t=-e_B+i_PR_P \tag{20}$$

$$\Sigma mmf_B=0=-N_Pi_P+N_CI_C+N_Ci_C+N_Li_L \tag{21}$$

$$e_B=-\frac{N_P}{N_L}i_LR_L=-\frac{N_P}{N_C}i_CR_C \tag{22}$$

These equations will yield the following:

$$i_P=\frac{1}{1+\alpha}\left(\frac{\pi}{2}\frac{E_P}{R_P}\sin\omega t-\frac{N_C}{N_P}I_C\right) \tag{23}$$

$$i_L=\frac{1}{1+\alpha}\frac{N_L}{N_P}\left(\frac{\pi}{2}\frac{E_P}{R_L}\sin\omega t-\frac{N_C}{N_P}I_C\frac{R_P}{R_L}\right) \tag{24}$$

$$e_B=-\frac{1}{1+\alpha}\left(\frac{\pi}{2}E_P\sin\omega t-\frac{N_C}{N_P}I_CR_P\right) \tag{25}$$

Interval 3 starts at θ_3 when core A saturates: interval 4 starts at θ_4 when core B saturates; interval 4 is the time from which both curves are in saturation until core B comes out. When this happens, core B performs in the same way as did core A in interval 1. The process is thus repeated. Both θ_3 and θ_4 must be determined from an examination of the flux excursions in the cores.

The flux excursion of core A from negative to positive saturation occurs during intervals 1 and 2. The volt-seconds applied to core A during these intervals can be found by integrating equations 8 and 16 with respect to time. This integral will be equal to the volt-seconds applied by the saturation voltage for one half-cycle.

$$\pi E_S=\frac{\pi}{2}E_P\left(\frac{\cos\theta_1-\cos\theta_2}{1+\alpha}+\frac{\cos\theta_2-\cos\theta_3}{2}\right)+\frac{N_C}{N_P}I_CR_P\times\left(\frac{\theta_2-\theta_1}{1+\alpha}+\frac{\theta_3-\theta_2}{2\alpha}\right) \tag{26}$$

With the use of the excitation factor $\bar{x}$ and equation 18, expression 26 can be simplified to become

$$\frac{2}{\bar{x}}=\frac{\cos\theta_1-\cos\theta_2}{1+\alpha}+\frac{\cos\theta_2-\cos\theta_3}{2}+\sin\theta_2\left(\frac{\alpha}{1+\alpha}(\theta_2-\theta_1)+\frac{\theta_3-\theta_2}{2}\right) \tag{27}$$

The flux excursion of core B from negative to positive saturation takes place during intervals 2 and 3. Integrating equations 17 and 25 and equating to the saturation volt-second integral yields the following relationship:

$$\frac{2}{\bar{x}}=\frac{\cos\theta_2-\cos\theta_3}{2}+\frac{\cos\theta_3-\cos\theta_4}{1+\alpha}-\sin\theta_2\left[\frac{\theta_3-\theta_2}{2}+\frac{\alpha}{1+\alpha}(\theta_4-\theta_3)\right] \tag{28}$$

The average value of the negative half of the load current is found by integrating equations 7 and 15 over a half-cycle. This integration yields

$$I_L = -\frac{1}{2}\frac{N_L}{N_P}\frac{E_P}{R_L}\times \left[\frac{\cos\theta_1-\cos\theta_2}{1+\alpha}+\frac{N_CI_C}{N_P\frac{\pi}{2}\frac{E_P}{R_P}}\times \left(\frac{\theta_2-\theta_1}{1+\alpha}+\frac{\theta_3-\theta_2}{\alpha}\right)\right] \tag{29}$$

Substitution of equations 18 and 27 into equation 29 will yield

$$I_L = -\frac{1}{4}\frac{N_L}{N_P}\frac{E_P}{R_L}\times \left[\frac{4}{\bar{x}}-(\cos\theta_2-\cos\theta_3)+(\sin\ \theta_2)(\theta_3-\theta_2)\right] \tag{30}$$

The average value of the positive half of the load current is found by integrating equation 24:

$$I_L = \frac{1}{2}\frac{1}{1+\alpha}\frac{N_L}{N_P}\frac{E_P}{R_L}\times \left[\cos\theta_3-\cos\theta_4-\frac{N_CI_C}{N_P\frac{\pi}{2}\frac{E_P}{R_P}}(\theta_4-\theta_3)\right] \tag{31}$$

Substituting equations 18 and 28 into equation 31 shows that this current is

$$I_L = \frac{1}{4}\frac{N_L}{N_P}\frac{E_P}{R_L}\times \left[\frac{4}{\bar{x}}-(\cos\theta_2-\cos\theta_3)+(\sin\ \theta_2)(\theta_3-\vartheta_2)\right] \tag{32}$$

The average value of the positive half is equal to that of the negative half. This means that the average load current over each half-cycle is zero, i.e., there is no d-c component present.

The following per-unit quantities are defined:

$$\bar{i}_L = \frac{I_L}{\frac{N_L}{N_P}\frac{E_S}{R_L}} \tag{33}$$

$$\bar{a} = \frac{N_CI_C}{N_L\frac{N_L}{N_P}\frac{E_S}{R_L}} \tag{34}$$

$$\bar{a} = \frac{\pi}{2}\bar{x}\left(1+\frac{N_C^2}{N_L^2}\frac{R_L}{R_C}\right)\sin\theta_2 \tag{35}$$

The per-unit average value of each half of the load current as defined by equations 32 and 33 is

$$\bar{i}_L = \frac{\bar{x}}{4}\left[\frac{4}{\bar{x}}-(\cos\theta_2-\cos\theta_3)+(\theta_3-\theta_2)\sin\theta_2\right] \tag{36}$$

Equation 27 was solved by a graphical method. The relationship between θ_1 and θ_2 is known from equation 9 and 18. Various values of θ_2 were assumed. Then, by assuming different values of θ_3, for each value of θ_2, a value of $\bar{x}$ for each θ_2, θ_3 combination was determined. From these data, a family of curves of θ_3 versus $\bar{x}$ for the various values of θ_2 were plotted. A typical plot is shown in Fig. 7. Then, by choosing a value of $\bar{x}$, the parameter θ_2 and θ_3 can be taken from the plot and substituted in equation 18 to determine the per-unit load current. In this manner the data from the curves shown in Fig. 4 were obtained. The values of $\bar{x}=4$ and $\bar{x}=6$ were chosen to illustrate the effect of varying line voltage.

When θ_2 is equal to θ_3, the flux excursion in core *A* from one saturated condition to the other occurs before core *B* comes out of saturation. Under this condition, the maximum possible voltage will be transferred to the load. Substituting this condition into equation 32 yields

$$I_{L,\max} = \frac{N_L}{N_P}\frac{E_S}{R_L} \tag{37}$$

The power output of the modulator will be:

$$P_0 = I_LR_L^2 = \left(\frac{N_L}{N_P}\right)^2\frac{E_S^2}{R_L}\left(\frac{\bar{x}}{4}\right)^2\times \left[\frac{4}{\bar{x}}-(\cos\theta_2-\cos\theta_3)+(\theta_3-\theta_2)\sin\theta_2\right]^2 \tag{38}$$

The maximum power output will be

$$P_{0,M} = \left(\frac{N_L}{N_P}\right)^2\frac{E_S^2}{R_L} \tag{39}$$

For extremely low-level signals, as in

$$I_C \ll \frac{\pi}{2}\frac{N_P}{N_C}\frac{E_P}{R_P} \tag{40}$$

and

$$I_C \ll \frac{1}{\alpha}\frac{\pi}{2}\frac{N_P}{N_C}\frac{E_P}{R_P} \tag{41}$$

θ_1 and θ_2 are essentially zero, and equations 27 and 36 reduce to

$$\frac{\bar{a}}{\bar{x}}\left(\frac{1}{1+\left(\frac{N_C}{N_L}\right)^2\frac{R_L}{R_L}}\right) = \frac{\frac{4}{\bar{x}}-1+\cos\theta_3}{\frac{2\theta_3}{\pi}} \tag{42}$$

$$\bar{i}_L = \frac{x}{2}\left(\frac{4}{\bar{x}}-1+\cos\theta_3\right) \tag{43}$$

For this case, the per-unit load current can be expressed as

$$\bar{i}_L = \bar{a}\left(\frac{1}{1+\left(\frac{N_C}{N_L}\right)^2\frac{R_L}{R_C}}\frac{\theta_3}{\pi}\right) \tag{44}$$

Both cores are driven from one value of saturation to the other during each half-cycle of supply frequency. Thus, if a step change in current is applied to the control winding, it will cause a transient change in the load current in the half-cycle in which it is applied. On the subsequent half-cycle, the cores will start from a saturated condition. The output current will be at its new quiescent value. Thus, this circuit can be said to have half-cycle response.

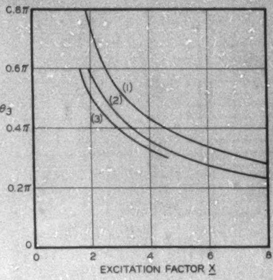


Fig. 7. Phase angle θ_3 versus excitation factor for $\alpha=0.1$

Curve 1—$\theta_2=0.05\ \pi$
Curve 2—$\theta_2=0.15\ \pi$
Curve 3—$\theta_2=0.25\ \pi$

Nomenclature

$\bar{a}$ = per-unit control ampere turns
$e_{A,B}$ = instantaneous voltage across primary winding of cores *A* and *B*, volts
E_P = average value of supply voltage, volts
i_L = instantaneous current in load, circuit, amp (amperes)
i_P = instantaneous current in primary circuit, amp
i_C = a-c component of control current, amp
I_C = d-c component of control current, amp
I_L = average value of each half-cycle of load current, amp
$\bar{i}_L$ = per-unit load current
N_L, N_P, N_C = number of turns on each core of the load, primary and control windings respectively
R_L = total load circuit resistance, ohms
R_C = total control circuit resistance, ohms
R_p = total primary circuit resistances, ohms
$\bar{x}$ = per-unit excitation factor, E_p/E_s
α = reflected resistance ratio
θ_1 = phase angle at which first core comes out of saturation, radians
θ_2 = phase angle at which second core come out of saturation, radians
θ_3 = phase angle at which first core goes into saturation at opposite polarity, radians
θ_4 = phase angle at which second core goes into saturation at opposite polarity, radians

References

1. THE FUNDAMENTAL LIMITATIONS OF THE SECOND-HARMONIC TYPE OF MAGNETIC MODULATOR AS APPLIED TO THE APPLICATION OF SMALL D. C. SIGNALS, **F. C. Williams, S. W. Noble.** *Proceedings*, Institution of Electrical Engineers, London, England, vol. 97, pt. II, 1950, pp. 445–59.

2. THE STUDY OF A MAGNETIC INVERTER FOR AMPLIFICATION OF LOW-INPUT-POWER D. C. SIGNALS, **E. H. Frost-Smith.** *Ibid.*, vol. 100, pt. II, 1953, pp. 362–70.

3. MAGNETIC AMPLIFIERS (book), **H. F. Storm.** John Wiley & Sons, Inc., New York, N. Y., 1955.

Inductors Biased with Permanent Magnets

Part I—Theory and Analysis

J. T. LUDWIG
ASSOCIATE MEMBER AIEE

IT HAS LONG been known that when a highly unsymmetrical ($I_{\min} \neq I_{\max}$) current magnetizes a closed iron core in an inductor or transformer, the optimum design includes an air gap in the core to reduce saturation. The presence of the air gap has two opposing effects: 1. it increases the reluctance of the core to both the biasing d-c magnetization and the incremental a-c magnetization; 2. the shift of magnetic operating point due to effect 1 reduces the reluctance of the iron to the incremental magnetization. The result of these two opposing effects is that as an air gap is increased from zero length in an iron core with unsymmetrical magnetization, the inductance of a coil linking the core first increases as the second effect is dominant and then a maximum is reached, after which the inductance decreases as the first effect is dominant.

If a permanent-magnet slab is used as a counterbiasing means instead of an air gap, the magnetic operating point may be shifted toward the maximum incremental permeability point with a gap much shorter than the optimum air gap. The direct increase in reluctance is then much smaller than in the air gap case, and the inductance is much larger. A given core size can supply a larger inductance or, conversely, a smaller permanent-magnet-gapped core can supply the same inductance as a larger air-gapped core.

Because of the thin gaps involved, it is not practical to laminate metal permanent magnets to form the gaps. If they were not laminated, eddy currents would nullify the advantages. The advent of barium ferrite,[1,2] a high resistivity permanent magnet, has made such counterbiasing gaps practical.[3–7] Such material can be inserted as thin slabs in gaps in ceramic or laminated metallic cores, or it can be formed directly as thin layers within ceramic cores.

Methods of air core inductor analysis have been discussed by several writers[8–22] with various degrees of approximation and trial-and-error procedure. The analysis method given by Karapetoff[8] in 1911 for finding the induction in a series magnetic circuit when a known d-c magnetomotive force is applied is of particular significance. This method involves an air gap line superposed on the iron normal induction curve. It requires no trial-and-error procedures, in contrast the usual method which requires the assumption of an induction and the calculation of the resulting magnetomotive force for comparison with the given magnetomotive force. Ludwig[3,22] has shown how the Karapetoff method may be extended, by the use of various induction curves, for the analysis of air-gapped inductors with superposed a-c and d-c (unsymmetrical) magnetization. The same general principle has had its application further extended to the analysis of biased inductors. A method was developed for analysis of an inductor with either specified current or specified voltage.

A synthesis method for biased inductors is presented in Part II of this report.[23]

The Basic Equations

An "ideal" magnetic circuit will be considered in the mathematical analysis. The magnetic induction is constant around the circuit, and there is no fringing of flux lines at points where the permeability changes. Because of the linear nature of the gap materials considered here, the corrections for fringing (spatial variation of induction) may be made almost independent of the main problem.

The fundamental equations relating the electrical conditions in the coil to the magnetic conditions in the core will be considered.

The core comprises a magnetically soft iron circuit interrupted by a magnetically hard gap. Both these ferromagnetic materials have inductions not proportional to their magnetizing forces. The appropriate *bh* relationships are introduced into the analysis by the following means. It should be noted that a mixed system of units is used. All the magnetic and spatial units are in the cgs (centimeter-gram-second) electromagnetic system, while the electrical units are in the mks (meter-kilogram-second) practical system. This mixed system is used because in United States practice nearly all electrical specifications are in mks units and nearly all magnetic material data are in cgs electromagnetic units.

In the iron, the *bh* relationship is very complicated. It is neither linear nor single-valued, and it is this complicated relationship that makes the design of magnetic circuits a difficult and esoteric art.

The first relation between the induction b and the magnetizing force h is the normal induction curve and the associated normal permeability. The normal induction curve is the locus of the tips of all symmetrical hysteresis loops. The coordinates of any point on the normal induction curve are then the maximum values, B_m and H_{mi}, for some symmetrical loop. The normal permeability μ_i for any point on the normal induction curve is defined by

$$\mu_i \overset{\Delta}{=} \frac{B_m}{H_{Mi}} \tag{1}$$

Since μ_i is a function of either B_m or H_{mi}, the choice of the independent variable is arbitrary. In the method of analysis, it is more convenient to use the direct relation between B_m and H_{mi} (the normal induction curve), but in the synthesis method it is more convenient to use the normal permeability.

A second relation between b and h is necessary to describe unsymmetrical (minor) hysteresis loops in which the maximum and minimum values are not the equal negatives of each other. The maximum point of a minor loop lies on the normal induction curve within engineering accuracy.[25–28]

An incremental permeability is associated with each unsymmetrical or minor loop. It is usually defined as

$$\mu_{\Delta i} = \frac{B_\Delta}{H_{\Delta i}} \tag{2}$$

where

B_Δ = half the total change of induction in gausses and is called the incremental induction

$H_{\Delta i}$ = half the total change of magnetizing force in oersteds and is called the incremental magnetizing force

The incremental permeability represents the slope of the minor loop and is a

---

Paper 60-198, recommended by the AIEE Basic Sciences Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., January 31–February 5, 1960. Manuscript submitted November 24, 1958; made available for printing December 11, 1959.

J. T. LUDWIG is with Electronic Communications, Inc., St. Petersburg, Fla.

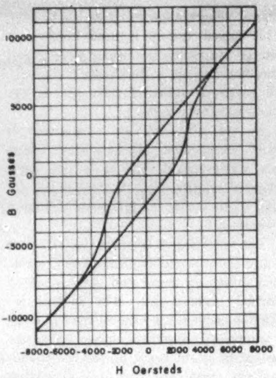


Fig. 1. Hysteresis loop of barium ferrite

function of any two variables which specify the tips of the minor loop.

The relative location of the minor loop is determined by the upper tip which falls on the normal induction curve so the ratio of B_m to H_{mi} is the normal permeability. Another method of specifying the location of the minor loop is by means of the center point whose co-ordinates are B_b and H_{bi}, the biased induction and the biasing magnetizing force, respectively. The relationship between B_b and H_{bi} depends on B_Δ or $H_{\Delta i}$ as well as B_b or H_{bi}. Thus, the use of this method of locating the minor loop involves one more variable and is, therefore, not used in this development except where data are already presented in terms of B_b or H_{bi}.

Since the magnetic induction is uniform around the circuit, the B's are not doubly subscripted, and the development will make use of the permeabilities as functions of the inductions rather than functions of magnetizing forces. In this manner, μ_i will represent $\mu_i(B_m)$ where B_m is the peak induction of a major (symmetrical) or minor (unsymmetrical) loop, and $\mu_{\Delta i}$ will represent $\mu_{\Delta i}(B_m, B_\Delta)$.

In the ferrite gap material, the bh relationship, like that of iron, is neither linear nor single-valued, but the scale is quite different. Fig. 1 shows the saturation hysteresis loop of unoriented barium ferrite. The portions of the loop utilized in this type of magnetic circuitry are substantially linear.

In a magnetic circuit containing no magnetomotive force sources, a permanent magnet operates in the second or fourth quadrant of its hysteresis loop. Where a magnet is being used to depolarize a magnetic circuit containing magnetomotive force sources, the operating point may move slightly counterclockwise around the loop to fall in the third or first quadrants.

If polarities are chosen so that the premagnetization of the permanent-magnet gap material is negative induction and magnetizing force (third quadrant), then the equation of the loop in and near the fourth quadrant is

$$H_g = H_c + (H_c B/B_r) \qquad (3)$$

where

H_g = the magnetizing force in the gap in oersteds
H_c = the absolute value of the coercive force in oersteds (H axis intercept)
B_r = the absolute value of the residual induction in gausses (negative B axis intercept)

For the case of peak positive magnetization, equation 3 becomes

$$H_{mg} = H_c + (H_c/B_r)B_m \qquad (4)$$

To effect a better match of characteristics between the core and gap materials, the gap material may be placed at an angle other than 90 degrees with the main flux path. If the gap material is sufficiently thin, the flux will cross the gap perpendicular to it, and the induction (flux density) in the gap material will be reduced by the factor sin θ. Since the actual gap material characteristics enter the equations by means of H_c and B_r, an equation for the effective residual induction may be written as

$$B_r = (B_{ra}/\sin\theta) \qquad (5)$$

where

B_r = the effective residual induction in gausses
B_{ra} = the actual residual induction in the gap material in gausses

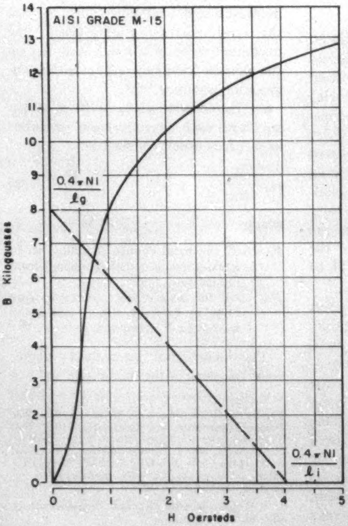


Fig. 2 (left). The normal induction curve with Karapetoff's air gap line to find d-c induction in absence of a-c induction

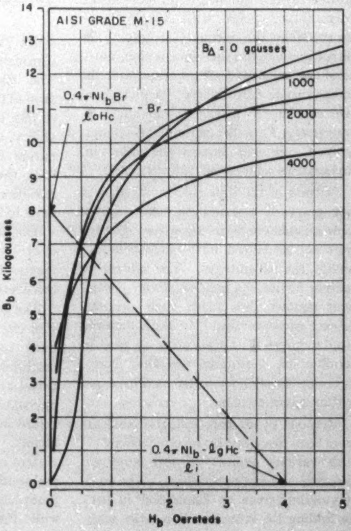


Fig. 3 (right). The biased (anhysteretic) induction curves with gap line to find the center of the minor loop

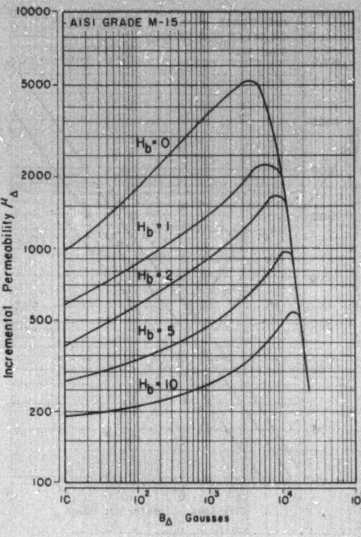


Fig. 4 (left). Incremental permeability curves

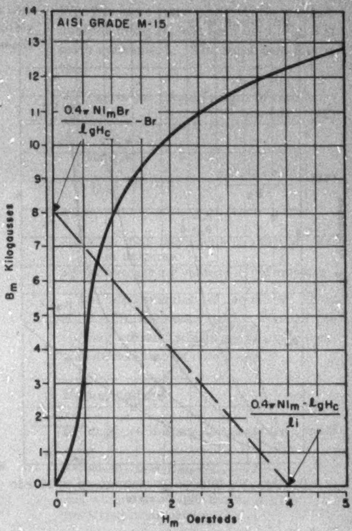


Fig. 5 (right). The normal induction curve with gap line to find the upper tip of the minor loop

The numbers for coercive force H_c and gap length ℓ_g do not need correction for angular gaps. Equation 5 for B_r in terms of B_{ra} will not be introduced directly into the synthesis method since it can be treated separately.

From the basic magnetomotive force equation, the following equations may be derived:

$$B_m = \frac{0.4\pi NI_m - \ell_i \beta H_c}{\ell_i[(1/\mu_i) + (\beta H_c/B_r)]} \qquad (6)$$

$$B = \frac{0.4\pi NI_\Delta}{\ell_i[(1/\mu_{\Delta i}) + (\beta/\mu_{\Delta g})]} \qquad (7)$$

which are basic equations for the design method.

Equations for the inductance of a coil linking this magnetic circuit may be derived from the definition of inductance:

$$L = \frac{NB_\Delta a_i}{I_\Delta 10^8} \qquad (8)$$

and

$$L = \frac{0.4\pi N^2 a_i 10^{-8}}{\ell_i[(1/\mu_{\Delta i}) + (\beta/\mu_{\Delta g})]} \qquad (9)$$

These are the basic inductance equations to be used in the analysis and in the synthesis method of the companion paper.[23]

The Analysis of Inductors

A method of analysis, determining the electrical characteristics from the physical configuration, is helpful to the design procedure in providing a check on designs and in evaluating the effect of changes of the various physical dimensions.

Because of the nonlinearity of the iron characteristics, the inductance of any iron-core inductor is not a constant but depends upon the electrical conditions under which it is operated. Methods of analysis will be developed for two types of electrical conditions:

1. The biasing current, I_b, and a time-varying component of voltage $e'(t)$, which produces a symmetrical time-varying current component, are specified.
2. The current $i(t)$ is specified.

ANALYSIS WITH THE SPECIFIED ALTERNATING VOLTAGE COMPONENT

When the time-varying component of voltage is given, the voltage drop across the resistance R_c is assumed to be small compared to the back electromotive force of the inductance.

By the usual methods,

$$B = \frac{\int_{t_1}^{t_2} e'dt 10^8}{2Na_i} \qquad (10)$$

where t_1 and t_2 are points of minimum and maximum $\int e'dt$, respectively.

When e' is a sinusoidal wave with frequency f and effective value E_{ac},

$$B_\Delta = \frac{E_{ac} \times 10^8}{\sqrt{2}\pi f N a_i} \qquad (11)$$

or

$$B_\Delta = \frac{E_{ac} \times 10^8}{4.44 f N a_i} \qquad (12)$$

which is the customary transformer equation.

Karapetoff[8] describes a method for finding the induction in a uniform magnetic circuit with air gaps when a known d-c magnetomotive force is applied. First it is assumed that the relation between B and H in the iron is known and plotted as in Fig. 2. If the circuit is magnetized by a magnetomotive force $0.4\pi NI$, the magnetomotive force equation is

$$0.4\pi NI = H\ell_i + B\ell_g \qquad (13)$$

or, alternatively,

$$\frac{0.4\pi NI}{\ell_i} = H + B\frac{\ell_g}{\ell_i} \qquad (14)$$

This is the equation of a straight line on the BH plane having intercepts

$$H = \frac{0.4\pi NI}{\ell_i} \qquad (15)$$

and

$$B = \frac{0.4\pi NI}{\ell_g} \qquad (16)$$

This straight line may be plotted by means of these intercepts or any two points which satisfy equation 14. Such a line is shown in Fig. 2. The intersection of this line with the iron BH curve then gives the induction and magnetizing force in the iron. The B value of this point is

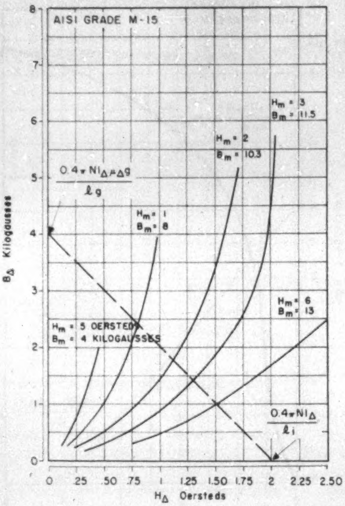


Fig. 6 (left). The incremental induction curves with gap line to find the incremental induction

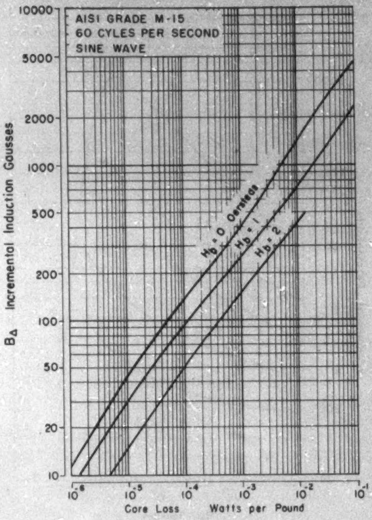


Fig. 7 (right). Core loss with biasing magnetizing force as a parameter

the induction and magnetizing force in the air gap.

This method can be extended to the problem of magnetic circuits with "hard" gaps. In such a circuit the magnetomotive force equation is

$$0.4\pi NI = H\ell_i + \ell_g[(BH_c/B_r)+H_c] \tag{17}$$

Dividing by ℓ_i and rearranging gives

$$\frac{0.4\pi NI}{\ell_i} - H_c\frac{\ell_g}{\ell_i} = H + \frac{\ell_g}{\ell_i}\frac{H_c}{B_r}B \tag{18}$$

This is the equation of a straight line with intercepts

$$H = \frac{0.4\pi NI}{\ell_i} - H_c\frac{\ell_g}{\ell_i} \tag{19}$$

$$B = \frac{0.4\pi NI}{\ell_g}\frac{B_r}{H_c} - B_r \tag{20}$$

When this line is plotted together with the appropriate BH curve, their intersection is the solution.

A comparison of these hard gap lines with the air gap lines shows that for the case of the hard gap, both intercepts are closer to the origin than in the case of the air gap. Both of these effects tend to move the operating point (intersection) toward the origin with a subsequent increase in incremental permeability and inductance for a given structure.

In the analysis of the inductor with specified I_b and $e'(t)$, the extended method can be applied to find the B_b, H_b point.[29] A set of plots of B_b versus H_b for various B_Δ values is shown in Fig. 3. The equation of the Karapetoff line is

$$\frac{0.4\pi NI_b}{\ell_i} - H_c\frac{\ell_g}{\ell_i} = H_b + \frac{\ell_g}{\ell_i}\frac{H_c}{B_r}B_b \tag{21}$$

and its intercepts are

$$H_b = \frac{0.4\pi NI_b - \ell_g H_c}{\ell_i} \tag{22}$$

and

$$B_b = \frac{0.4\pi NI_b}{\ell_g}\frac{B_r}{H_c} - B_r \tag{23}$$

The values of B_b and H_b in the iron may be found at the intersection of this line with the B_b H_b curve for the particular value of B_Δ, already found by equation 10 or 12.

One method of presenting incremental permeability data is shown in Fig. 4. Several curves show μ_Δ as a function of B_Δ, each for a different value of H_b. The values of B_Δ and H_b already found are used to find $\mu_{\Delta i}$ directly from this graph.

The inductance L may be calculated by equation 9:

$$L = \frac{0.4\pi N^2 a_i 10^{-8}}{\ell_i[(1/\mu_{\Delta i})+(\beta/\mu_{\Delta g})]}$$

Thus, when an alternating voltage and a direct current are specified, the analysis steps are:

1. Calculate B_Δ from equation 10 or 12.
2. Construct Karapetoff line on the biased induction curves (B_b vs H_b). For the value of B_Δ found above, read B_b and H_b of equations 21, 22, and 23.
3. For B_Δ and H_b found above, find $\mu_{\Delta i}$ from curve, Fig. 4.
4. Calculate the inductance from equation 9.

Analysis with Specified Current

When the current, $i(t)$, is specified, a different analysis method is used. First, the minimum and maximum values of $i(t)$ are determined and I_Δ is calculated by

$$I_\Delta = \frac{I_m - I_{\min}}{2} \tag{24}$$

The extended Karapetoff-type method is applied to the normal induction curve such as Fig. 5. The equation of the straight line is

$$\frac{0.4\pi NI_m}{\ell_i} - \frac{\ell_g H_c}{\ell_i} = H_m + \frac{\ell_g}{\ell_i}\frac{H_c}{B_r}B \tag{25}$$

and its intercepts are

$$H_m = \frac{0.4\pi NI_m - \ell_g H_c}{\ell_i} \tag{26}$$

and

$$B_m = \frac{0.4\pi NI_m}{\ell_g}\frac{B_r}{H_c} - B_r \tag{27}$$

The intersection of the normal induction curve and the straight line is the H_m, B_m point of the minor loop, because the right hand (H_m, B_m) tips of all the minor loops lie on the normal induction curve independently of B_Δ or H_Δ.

Instead of considering the incremental permeability curves as before, the incremental induction (B_Δ vs H_Δ) curves

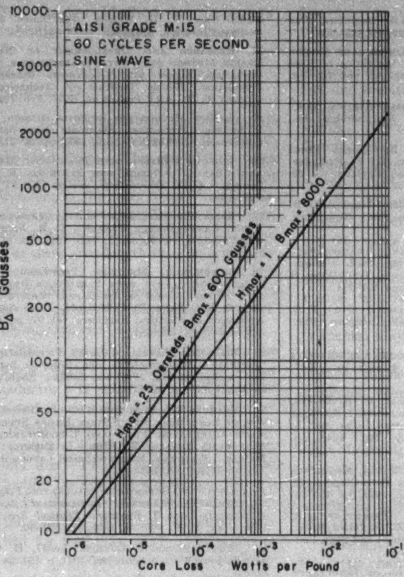


Fig. 8. Core loss with maximum magnetizing force as a parameter

shown in Fig. 6 will be used with a Karapetoff line construction. Each incremental induction curve is associated with one (H_m, B_m) point on the normal induction curve, and the appropriate curve (or interpolated curve) is used for the (H_m, B_m) value found by the extended Karapetoff-type line construction on the normal induction curve. The line on this incremental induction graph is drawn to fit the equation

$$\frac{0.4\pi NI_\Delta}{\ell_i}=H_\Delta+\frac{\ell_g}{\ell_i}\frac{B_\Delta}{\mu_{\Delta g}} \qquad (28)$$

which may be derived from Ampere's law equations for I_m and I_Δ points. The intercepts are

$$H_\Delta=\frac{0.4\pi NI_\Delta}{\ell_i} \qquad (29)$$

and

$$B_\Delta=\frac{0.4\pi NI_\Delta\mu_{\Delta g}}{\ell_i} \qquad (30)$$

The intersection of this line with the proper incremental induction curve gives H_Δ and B_Δ in the iron, and the B_Δ value is also the incremental induction and incremental magnetizing force in the air gap. The inductance, L, can now be found from equation 8:

$$L=\frac{NB_\Delta a_i}{I_\Delta}10^{-8}$$

Thus, when the current is specified, the analysis steps are:

1. Construct hard gap line on normal induction curve by equations 25, 26, and 27. Read B_m and H_m.
2. Construct a Karapetoff-type line on incremental induction curves by equations 28, 29, and 30. For the value of B_m found above, read B and H.
3. Calculate the inductance from equation 8.

ANALYSIS OF RESISTANCE

In either case, specified voltage or specified current, the copper resistance R_c is calculated by the usual methods. The general resistance equation is

$$R_c=(\rho N\ell_c/a_w) \qquad (31)$$

where

ρ = the resistivity in ohm-centimeters
N = the number of turns
ℓ_c = the mean length of turn in centimeters
a_w = the cross-sectional area of the wire in square centimeters

A more convenient method for the resistance calculation is the use of a wire table for the value of ρ/a_w and the appropriate $N\ell_c$ in the units consistent with the wire table units.

The core loss component of resistance is calculated by use of incremental core loss data as in Fig. 7 or Fig. 8. If the voltage analysis using H_b has been used, Fig. 7 is more convenient; if the current analysis using H_m has been used, Fig. 8 is more convenient. The total core loss is then

W_i = watts per pound (weight of core in pounds) (32)

and

$$R_i=\frac{W_i}{I_{rms}^2-I^2} \qquad (33)$$

where

I_{rms} = the root-mean-square value of $i(t)$ in amperes
I = the average d-c value of $i(t)$ in amperes

The denominator of equation 33 is ($I_{rms}^2-I^2$), the mean square value of the varying current which is the only component resulting in core losses.

Nomenclature

Subscripts on time variable quantities (period P)

b = biased, $Z_b=(Z_{max}+Z_{min})/2$
m = maximum value during a cycle
min = minimum value during a cycle
rms = root-mean-square, $Z_{rms}=\sqrt{(1/P)\int_0^P z^2dt}$
Δ = incremental, $Z=(Z_{max}-Z_{min})/2$
no subscript = time average, $Z=(1/P)\int_0^P zdt$

General Subscripts

g = value within the gap
i = value within the iron

Letter Symbols

a_i = net iron cross-section area in cm^2
a_w = net wire cross-sectional area in cm^2
b = time variable induction in gausses
B_r = gap material effective residual induction in gausses
B_{ra} = gap material actual residual induction in gausses
e' = a time variable voltage in volts which produces a symmetrical time varying current
E_{ac} = sine wave effective (rms) voltage in volts
f = frequency in cycles per second
h = time variable magnetizing force in oersteds
H_c = gap material coercive force in oersteds
i = time variable current in amperes
ℓ = generalized length in centimeters
ℓ_c = mean length of turn in centimeters
ℓ_g = effective length of gap in centimeters
ℓ_i = mean iron length in centimeters
L = inductance in henrys
N = number of turns
R_c = copper resistance in ohms
R_i = core loss component of resistance in ohms
t = time in seconds
W_i = iron core loss in watts

Greek Symbols

β = ratio of effective gap length to iron length, ℓ_g/ℓ_i
θ = angle between the flux lines and the plane of the gap
μ = normal permeability in gausses
μ_Δ = incremental permeability in gausses per oersted

μ_t = normal permeability of the iron (distinguished from the gap)

ρ = coil resistivity in ohm-centimeters

References

1. Ferroxdure. A Class of New Permanent Magnet Materials, **J. J. Went, G. W. Rathenau, E. W. Gorter, G. W. VanOosterhout.** *Philips Technical Review*, Eindhoven, Netherlands, vol. 13, Jan. 1952, pp. 28–37.

2. Ferroxdure II and III, Anisotropic Permanent Magnet Materials, **A. L. Stuijts, G. W. Rathenau, G. H. Weber.** *Ibid.*, vol. 16, Nov.-Dec., 1954, pp. 141–47.

3. Final Report, Research and Development on Design Method for Reactors, **J. T. Ludwig.** University of Minnesota, Minneapolis, Minn., Contract no. DA36-039-SC-42573, Sept. 15, 1954.

4. The Application of Ferroxdure for Premagnetizing Cores in Power Carrying Coils, *Matronics*, Philips Electrical Ltd., Eindhoven, Netherlands, vol. 9, June 1955, pp. 153–60.

5. Ignition Coils with Ferroxcube and Ferroxdure. *Ibid.*, pp. 161–64.

6. Premagnetization of the Core of a Pulse Transformer by Means of Ferroxdure, **H. G. Bruijning, A. Rademakers.** *Philips Technical Review*, vol. 19, July 1957, pp. 28–37.

7. Design of Optimum Inductors Using Magnetically Hard Ferrites in Combination with Magnetically Soft Materials, **J. T. Ludwig.** *Journal of Applied Physics*, New York, N. Y., vol. 29, no. 3, Mar. 1958, pp. 497–99.

8. The Magnetic Circuit (book), **V. Karapetoff.** McGraw-Hill Book Company, Inc., New York, N. Y., 1911, p. 29, problem 11.

9. Design Methods for Soft Magnetic Materials in Radio, **J. Minton, I. G. Maloff.** *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 17, June 1929, pp. 1021–33.

10. The Air Gap Transformer and Choke, **F. W. Lanchester.** *Journal*, Institution of Electrical Engineers, London, England, vol. 73, Oct. 1933, pp. 413–18.

11. Magnetic Core Materials Practice (book). Allegheny Steel Company, Brackenridge, Pa., pt. III, 1937, pp. 66–86.

12. The Inductance of Iron-Cored Coils Carrying Direct Current, **G. F. Partridge.** *Philosophical Magazine*, London, England, vol. 23, 1937, pp. 99–106.

13. Electromagnetic Devices (book), **H. E. Roters.** John Wiley & Sons, Inc., New York, N. Y., 1941, pp. 102–05.

14. Magnetic Circuits and Transformers (book), **Electrical Engineering Staff.** Massachusetts Institute of Technology, John Wiley & Sons, Inc., 1943, pp. 68–81, and 197–241.

15. Radio Engineers' Handbook (book), **F. E. Terman.** McGraw-Hill Book Company, Inc., 1943, pp. 97–107.

16. Electronic Transformers and Circuits (book), **R. Lee.** John Wiley & Sons, Inc., 1947, pp. 66–77.

17. Pulse Generators (book), **G. N. Glasoe, J. V. Lebacqz,** editors. Massachusetts Institute of Technology Radiation Laboratory Series, vol. 5, McGraw-Hill Book Company, Inc., 1948, pp. 372–80.

18. Reference Data for Radio Engineers (book). Federal Telephone and Radio Corporation, New York, N. Y., third edition, 1949, pp. 194–95.

19. Transformers, Their Principles and Design for Light Electrical Engineers (book), **F. C. Connelly.** Sir Isaac Pitman and Sons, London, England, 1950, pp. 405–15.

20. The Theory and Design of Inductance Coils (book), **V. G. Welsby.** MacDonald and Company, London, England, 1950, pp. 82–85.

21. Research and Development of New Design Method for Power Transformers, **H. Garbarino.** *Report no. 21*, Armour Research Foundation, Illinois Institute of Technology, Chicago 16, Ill., ARF Project no. E017, Feb. 1953.

22. A New Method for Inductor Design, **J. T. Ludwig.** *Proceedings*, Electronic Components Conference, Los Angeles, Calif., 1955, pp. 20–21.

23. Inductors Biased with Permanent Magnets: Part II—Design and Synthesis, **J. T. Ludwig.** *AIEE Transactions*, see pp. 278–91 of this issue.

24. Design of Reactances and Transformers Which Carry Direct Current, **C. R. Hanna.** *AIEE Journal*, vol. 46, no. 1, Feb. 1927, pp. 128–31.

25. Incremental Magnetic Properties of Silicon-Iron Alloys with Particular Reference to the Design of Air-Gapped Smoothing Chokes, **R. O. Carter, D. L. Richards.** *Proceedings*, Institution of Electrical Engineers, vol. 97, pt. II, Apr. 1950, pp. 199–214.

26. Magnetic Materials (book), **F. Brailsford.** "Methuen's Monographs," John Wiley & Sons, Inc., second edition, 1951; also London, England, 1948, vol. 11, pp. 10–12.

27. Further Study on the Magnetic Properties of Electrical Sheet Steel Under Superposed Alternating Field and Unsymmetrical Hysteresis Losses, **Y. Niwa, J. Sugiura, J. Matura.** *Researches*, Electrotechnical Laboratory, Tokyo, Japan, no. 144, p. 19.

28. Effect of a Superposed Alternating Field on Apparent Magnetic Permeability and Hysteresis Loss, **T. Spooner.** *Physical Review*, New York, N. Y., vol. 25, 1925, pp. 527–40.

29. Electromagnetic Devices (book), **H. C. Roters.** John Wiley & Sons, Inc., 1941, p. 104, footnote 6.

Inductors Biased with Permanent Magnets

Part II—Design and Synthesis

J. T. LUDWIG
ASSOCIATE MEMBER AIEE

INDUCTORS biased with permanent magnet slabs and their analysis have been discussed in Part I of this report.[1] The design of such devices is discussed here. It is desirable that a method of design proceed from the specified electrical characteristics directly to the physical dimensions of the inductor without the use of trial-and-error methods.

The literature of air-gapped inductor synthesis includes papers by Hanna,[2] Crowhurst,[3] Carter and Richards,[4] Legg,[5] and Ludwig.[6,7] Hanna's method,[2] which has widespread usage, will determine the best air gap to use for an inductor when the core size is known or assumed. Carter and Richards[4] present an algebraic method of synthesis and Crowhurst[3] presents a graphical method based on a similar derivation, but both of these assume that the maximum incremental permeability is a simple exponential function of the polarizing magnetizing force. This is not an entirely satisfactory approximation over the range normally encountered. Legg[5] has attacked the problem by solving the system of governing equations so a design chart may be prepared from the magnetic characteristics of the core material. This chart specifies all the physical dimensions when the electrical requirements are known.

The synthesis methods of Carter and Richards, Crowhurst, and Legg are all applied to the case of small a-c inductions and the use of specified copper resistance as the only size-limiting factor. The method of Ludwig[6,7] removed these limitations; it included the effects of both copper and iron losses in terms of both specified circuit resistance and thermal heating (specified surface power density). The basic attack on the problem was the simultaneous solution of the pertinent equations relating the electrical, magnetic, and physical factors in terms of the iron characteristics. With the electrical specifications known, the physical dimensions and magnetic factors may be found from charts obtained from the iron characteristics.

The present theory treats the permanent-magnet-biased inductor in an analogous manner. It also shows that the writer's previous theory and method for air-gapped inductors are a special limiting case of the present general theory and method for permanent-magnet-gapped inductors.

Paper 60-199, recommended by the AIEE Basic Sciences Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., January 31–February 5, 1960. Manuscript submitted June 29, 1959; made available for printing December 23, 1959.

J. T. Ludwig is with Electronic Communications, Inc., St. Petersburg, Fla.

The Design Method for Inductors Limited by Copper Resistance

SHAPE NUMBERS AND GENERAL EQUATIONS

In the mathematical formulation of the problem, it is convenient to separate size and shape. The spatial quantities of principal importance are: the mean iron length, the mean copper turn length, the net cross-sectional area of copper, and the net cross-sectional area of iron. The mean iron length may be arbitrarily chosen as a measure of size, and the other three quantities may be related to it by means of dimensionless multiplying factors u, v, and w respectively.

These numbers then describe the shape, while ℓ_t describes the size of the inductor. It should be noted that the v and w are defined in terms of the net cross-sectional area, which includes the respective space factors for the copper and iron. In the optimization process, finding the smallest size design, the u, v, and w are treated as constants to simplify the solution. This simplification is justified for two reasons: 1. It is relatively simple to make several designs with this constant-shape method to compare the advantage of different shapes for the particular specifications to be met. 2. The economic factors are strongly in favor of the "scrapless" *EI* lamination. A method of considering various stack heights will be described later.

By eliminating N between the general inductance equation 9 in the companion paper[1] and the resistance equation

$$R_c=u\rho L[(1/\mu_{\Delta t})+(\beta/\mu_{\Delta g})]10^8/[0.4\pi vw\ell_t^2] \tag{1}$$

Examination of equation 1 shows that the copper resistance varies directly with u, ρ, and $[(1/\mu_{\Delta t})+(\beta/\mu_{\Delta g})]$ and inversely with v, w, and ℓ_t^2 for a given inductance. The copper resistance is a dissipative factor which usually must be held under some limit. To make R_c small, the values of u, ρ, and $[(1/\mu_{\Delta t})+(\beta/\mu_{\Delta g})]$ should be as small as possible and the values of v and w should be as large as possible. Winding the conductor around the core with minimum insulation space will minimize u while the conductor material determines ρ. Minimum internal coil insulation and space-saving winding methods will maximize the copper space factor, F_c, and shape factor, v. Likewise, in the core, a high space factor, F_t, will maximize w.

The effective incremental reluctivity factor $(1/\mu_{\Delta t})+(\beta/\mu_{\Delta g})$ requires detailed consideration, because $\mu_{\Delta t}$ is a nonlinear function of the operating inductions. When an unsymmetrical magnetomotive force is applied to an air-gapped ferromagnetic core, a minimum value of effective incremental reluctivity may occur for a positive, nonzero value of air gap length. This occurs because of two conflicting factors. As the air gap length is increased from zero, it introduces incremental reluctance into the circuit, increasing the effective incremental reluctivity. It also, however, introduces normal reluctance which shifts the operating point closer to the origin. This shift of operating point decreases the iron incremental reluctivity more than enough to compensate for the direct increase in incremental reluctance due to the air gap. At some gap length, the two rates of change become equal; and there is a minimum value of effective incremental reluctivity.

In the case of a hard gapped core, the shift of operating point toward the origin is even greater for a given increase in direct gap incremental reluctance. This results in a deeper minimum value for the effective incremental reluctivity.

Since $\mu_{\Delta t}$ is a function of both B_m and B_Δ, two more fundamental equations are required.

$$B_m=[0.4\pi NI_m-\ell_t\beta H_c]/\{\ell_t[(1/\mu_t)+(\beta H_c/B_r)]\} \tag{2}$$

$$B_\Delta=[0.4\pi NI_\Delta]/\{\ell_t[(1/\mu_{\Delta t}+(\beta/\mu_{\Delta g})]\} \tag{3}$$

Equations 1, 2, and 3 are a system of 3 equations in 4 unknowns (β, B_m, B_Δ, ℓ_t) since μ_t and $\mu_{\Delta t}$ are functions of B_m and B_Δ. One more condition will generally be required to determine a single solution.

MINIMUM SIZE DESIGN

The first additional condition is that of minimizing $[(1/\mu_{\Delta t})+(\beta/\mu_{\Delta g})]$. The iron incremental permeability, $\mu_{\Delta t}$, is a function of two variables, B_m and B_Δ, but it is convenient to consider it as a function of B_m alone by fixing the ratio B_Δ/B_m. It is then necessary to consider various values of the B_Δ/B_m ratio. The incremental permeability treated in this way, as a function of B_m, will be symbolized by $\mu_{\Delta t}'$. It can be shown by appropriate algebra and calculus that the solution of equations 1, 2, and 3 with the minimum size condition is

$$\left\{(B_m+B_r)\left[B_mD_n-\mu_{\Delta g}\left(H_c+\frac{H_cB_m}{B_r}\right)D_\Delta\right]-\frac{B_r}{\mu}\right\}\times\sqrt[4]{\frac{1}{\mu_{\Delta t}'}+\frac{B_r}{H_c\mu_{\Delta g}}\left[B_mD_n-\mu_{\Delta g}\left(H_c+\frac{H_cB_m}{B_r}\right)D_\Delta-\frac{1}{\mu_t}\right]}=\sqrt[4]{n_1} \tag{4}$$

where

$$n_1=[(0.4\pi)^5v^3wI_m^4R_c^2 10^{-8}]/[u^3\rho^3L] \tag{5}$$

$$D_n=[d\mu_t/dB_m]/\mu_t^2 \tag{6}$$

and

$$D_\Delta=[d\mu_{\Delta t}'/dB_m]/\mu'_{\Delta t}{}^2 \tag{7}$$

Equation 4 relates B_m and the known factors of n_1, since D_N, D_Δ, μ_t, and $\mu_{\Delta t}$ are functions of B_m alone. Given an n_1, this equation specifies the optimum B_m, but it is an implicit relationship which cannot be solved directly. It can be solved in the opposite manner by assuming various values of B_m and tabulating the values of n_1. These points may then be plotted in a chart with B_m ordinates and n_1 abscissas. In practice the chart will be used the other way around; the quantity n_1 is an index function which may be calculated from the specifications to obtain the optimum magnetic operating point B_m.

Generally there is an upper limit beyond which no minimum of $[(1/\mu_{\Delta t}')+(\beta/\mu_{\Delta g})]$ exists. The curves will terminate at this limit. It is also necessary to obtain β, ℓ_t, and N to complete the design, and "application" function curves may be added to the chart for this purpose. The equation for β can be shown to be

$$\beta=\frac{B_r}{H_c}\left[D_nB_m-D_\Delta\mu_{\Delta g}\left(H_c+\frac{H_cB_m}{B_r}\right)-\frac{1}{\mu_t}\right]. \tag{8}$$

This quantity may be calculated and plotted on the chart against a second set of ordinates and the corresponding values of the abscissa n_1.

It can also be shown that

$$\ell_t=\sqrt{[(1/\mu_{\Delta t}')+(\beta/\mu_{\Delta g})]u\rho L10^8/[0.4\pi vwR_c]} \tag{9}$$

For computing ℓ_t, it is convenient to plot an application function g_1 which is defined as

$$g_1\triangleq\sqrt{[(1/\mu_{\Delta t}')+(\beta/\mu_{\Delta g})]} \tag{10}$$

so that

$$\ell_t=g_1\sqrt{u\rho10^8/(0.4\pi vw)}\sqrt{L/R_c} \tag{11}$$

The value of equation 11 may be found for each equation 8 β value and plotted against a third set of ordinates on the chart. From these points, ℓ_t may be calculated by equation 11.

The equation for N is

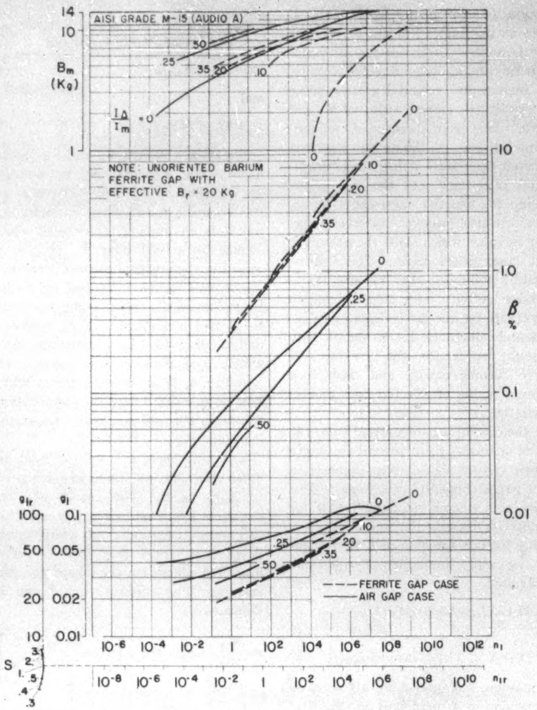


Fig. 1. Chart 1A: copper resistance limited designs, audio A

$$N=\sqrt{g_1\sqrt{v10^8/(0.4\pi u w \rho)}\sqrt{LR_c}} \qquad (12)$$

It should be recalled that the foregoing derivation was based on an additional restriction of a constant B_Δ/B_m ratio. Now the B_Δ/B_m ratio is not directly specified by the electrical requirements of the inductor so the method and chart developed above will not allow direct design from the electrical specifications. Several sets of calculations, each for a different value of B_Δ/B_m ratio, will be considered. The corresponding I_Δ/I_m ratio may be calculated from equations 2 and 3 for each point; from this information, design curves may be constructed in terms of the I_Δ/I_m ratio instead of the B_Δ/B_m ratio. These curves, each labeled with its respective I_Δ/I_m value, can then be used for direct design since the I_Δ/I_m ratio may be obtained from the electrical requirements.

Such a set of curves is presented in Chart 1A, Fig. 1. The dashed curves are for gaps of unoriented barium ferrite angled or arranged to have inductions one tenth those of the iron. The solid curves are for nonmagnetic gaps. Note that these curves, with index n_1 as abscissa and application functions B_m, β, and g_1 as ordinates, are determined solely by the magnetic characteristics of the core material, independent of the inductor geometry and the resistivity of the coil conductor.

To use the curves, the designer must calculate index n_1 by equation 5, and having found g_1 from the chart, he must calculate ℓ_i by equation 11 and N by equation 12.

AUXILIARY AXFS AND VARIABLE STACK HEIGHTS

If many designs are to be made with fixed constant values of u, v, w, and ρ, the design process may be simplified by adding two new axes to the chart. First consider the index function, n_{1r}, given by

$$n_{1r} \triangleq R_c^3 I_m^4/L = n_1 u^3 \rho^3 10^8/[(0.4\pi)^5 v^3 w] \qquad (13)$$

Since the n_1 axis is logarithmic, a logarithmic n_{1r} axis may be plotted parallel to it to perform the multiplication indicated in equation 13. The designer then need only calculate the middle term of equation 13 to find n_1 and enter the chart at this point on the n_{1r} axis. Unlike the n_1 axis, the n_{1r} axis depends on the shape numbers, u, v, and w and on the resistivity, ρ.

In a similar fashion, the calculations of ℓ_i from equation 11 can be simplified. Consider the application functions g_{1r} where

$$g_{1r} \triangleq \ell_i\sqrt{R_c/L} = g_1\sqrt{u\rho 10^8/[0.4\pi v w]} \qquad (14)$$

Having this value to insert in equation 11 simplifies the calculation. A logarithmic g_{1r} axis can be plotted parallel to the logarithmic g_1 axis to provide g_{1r} directly. The g_{1r} axis depends on u, v, w, and ρ.

From the g_1 or g_{1r} value obtained from the chart, the designer can calculate ℓ_i, the length of the core. This ℓ_i may have any value, while laminations are available only in discrete sizes, each with a discrete value of ℓ_i. One of the most convenient modifications of the standard shape is the variation of the height of the stack of laminations. In a qualitative manner, a lamination with an ℓ_i too small may be stacked higher to compensate; or, conversely, a lower stack of larger laminations may be used.

A variation in stack height will result in a change in u and w, since both the mean turn length and the net iron cross-section area are affected. If s is the ratio of the variable stack height to the standard stack height, the values of u and w will be symbolized generally by $u(s)$ and $w(s)$; and then $u(1)$ and $w(1)$ will be the standard values. The n_{1r} for the standard stack height axis is related to the n_1 axis by the factor $[u^3(1)\rho^3 10^8]/[(0.4\pi)^5 v^3 w(1)]$ and for variable stack heights the factor should be $[u^3(s)\rho^3 10^8]/[(0.4\pi)^5 v^3 w(s)]$. The n_{1r} axis for these variable stack heights should be shifted a distance of $\log\{[u^3(s)w(1)]/[u^3(1)w(s)]\}$.

By similar arguments, the g_{1r} axis for variable stack heights should be shifted a distance of $\log\{[u(s)w(1)]/[u(1)w(s)]\}^{1/2}$ from the position of the g_{1r} axis for standard stack height.

A convenient geometrical device for shifting the n_{1r} and g_{1r} axes is a transparent overlay with the n_{1r} and g_{1r} axes ruled on it. The main chart then may have an

Vol.
Pt. 1
196

auxiliary axis, s, to guide the location of some arbitrary reference point on the transparency. A convenient reference point is the intersection of the n_{1r} and g_{1r} axes. This auxiliary axis is the locus of points whose absicssas are $-\log\{[u^3(s)w(1)]/[u^3(1)w(s)]\}$ and whose ordinates are $-\log\{[u(s)w(1)]/[u(1)w(s)]\}$.

This curve can then be graduated with values of s. The transparency is placed over the chart with the corresponding axes parallel and with the intersection of the transparency axes lying on the s axis of the main chart. The n_{1r} and g_{1r} axes on the transparency will then be in correct relationship to the n_1 and g_1 axes and the curves on the chart for the particular value of s indicated on the s axis by the intersection of the n_{1r} and g_{1r} axes on the transparency. It should be noted that the application functions, B_m and β, are independent of the inductor geometry and are measured on the chart B_m and β axes directly. The abscissa (n_{1r}) for these curves must be the one on the transparency for stack heights other than standard.

The designer may use the transparency in either of two ways. If a design is wanted for a stack height other than standard, the transparency is located at the proper point on the s axis and the chart is read opposite the n_{1r} point on the transparency n_{1r} axis. The B_m and β values are read on the chart B_m and β axes, while the g_{1r} value is read on the transparency g_{1r} axis. If a design results in a value of ℓ_t intermediate between standard lamination sizes, the designer may determine the B_m, β, and s for the design using standard laminations having ℓ_t smaller or larger than that called for by the particular specifications.

He first calculates g_{1r} by equation 14 from the ℓ_t of the standard lamination and marks the n_{1r}, g_{1r} point on the transparency. He then moves the transparency, keeping the axes parallel to those on the main chart and keeping the intersection of the axes on the s axis of the main chart. When the n_{1r}, g_{1r} point on the transparency falls on the g_1 curve of the main chart, he can read s from the s axis at the intersection of the transparency n_{1r} and g_{1r} axes. The values of B_m and β are found from their respective curves and axes of the main chart, opposite the n_{1r} axis of the transparency.

This method of considering variation in stack height will be derived specifically for the case of "scrapless" *EI* laminations. For a stack of laminations s times as high as the center leg width

$$u(s)=\frac{2+s}{3} \qquad (15)$$

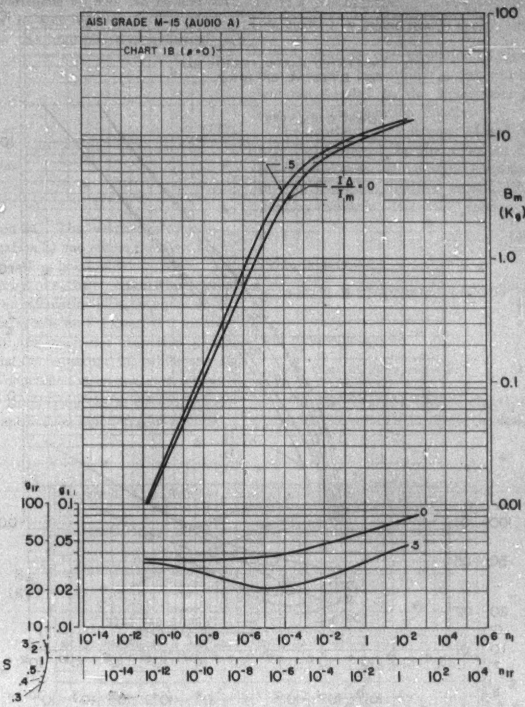


Fig. 2. Chart 1B: copper resistance limited designs with zero gap; audio A

$$v=0.00625 \text{ (for 30\% space factor)} \qquad (16)$$

and, for a 90% stacking factor,[8]

$$w(s)=(s/40)=0.025s \qquad (17)$$

The constant in equation 13 may now be evaluated for these values of u, v, w, and a copper resistivity of 1.724 microohm-centimeters:

$$n_{1r}/n_1=u^3\rho^3 10^8/(0.4\pi)^4 v^3 w=1/40 \qquad (18)$$

The constant in equation 14 is

$$g_{1r}/g_1=\sqrt{u\rho 10^8/0.4\pi v w}=1{,}000 \qquad (19)$$

The negative logarithmic co-ordinates of the points on the s axis are then $(2+s)^3/(27s)$ and $[(2+s)/(3s)]^{1/2}$. Such an axis is shown in Chart 1A, Fig. 1.

DESIGNS WITH ZERO AIR GAP

The method developed thus far provides a direct means of design when I_m, I_Δ, L, and R_c are specified. One limitation of this method is that there is no restriction on the sign of β; and for low values of n_1, the indicated gap length does become negative. For these specifications, the optimum design is the one with zero gap since a negative gap cannot be physically realized.

Previously the system of equations 1, 2, and 3 was solved in such a way that $(1/\mu_{\Delta t})+(\beta/\mu_{\Delta g})$ was a minimum. Now, for the zero gap case, the system of equations 1, 2, and 3 will be solved together with

$$\beta=0 \qquad (20)$$

The solution is

$$B_m{}^4/[\mu_t{}^4\mu_{\Delta t}{}']=n_1 \qquad (21)$$

Equation 21 gives the relation between B_m and the known requirements contained in n_1 since the μ's are functions of B_m. This equation is analogous to equation 4, and a design chart relating n_1 and B_m can be set up in the same way as Chart 1A was constructed. From equations 10 and 20

$$g_1=\sqrt{1/\mu_{\Delta t}{}'} \qquad (22)$$

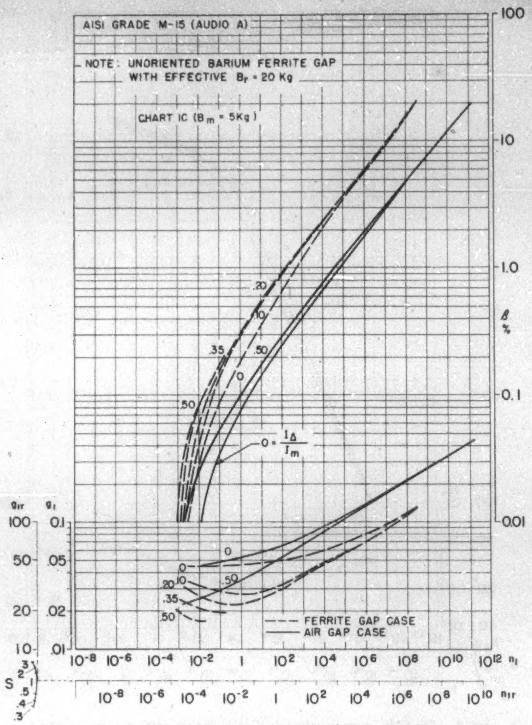


Fig. 3. Chart 1C: copper resistance limited designs with five kilogauss maximum induction; audio A

Equation 11 still applies for calculating ℓ_t, and equation 12 still applies for N.

This design chart has curves of application functions B_m and g_1 versus index n_1 for various I_Δ/I_m ratios. Axes for n_{1r}, g_{1r}, and s are identical to those of Chart 1A. This chart, for $\beta=0$, is shown as Chart 1B, Fig. 2.

Normally the designer would first calculate n_1 or n_{1r} and consult Chart 1A. If the value of n_1 or n_{1r} was too small to be found on Chart 1A, he would consult Chart 1B for the $\beta=0$ design which is the optimum design when the "minimum size" design gap is negative.

DESIGN WITH INDUCTION LIMITATION

Another limitation of the minimum size method of Chart 1A is that the B_m may be too large for the linearity or core loss requirements of the inductor. The specification and mensuration of linearity and nonlinearity is a very complicated question, especially when the waveforms of current and voltage are of unlimited shapes. The maximum induction B_m is a guide. If it is above the "knee" of the core material induction curve, minor loops with a large B_Δ will generally bend around the knee. If the peak induction B_m is great enough to drive the gap biasing material over the third quadrant corner, it will recoil with lower demagnetizing effect. It is then useful to have the design curves for various limiting values of B_m. If the limiting value of B_m is below the optimum value found in Chart 1A, the "B_m-limited" design will generally be larger. If the limiting value of B_m is above the optimum value found in Chart 1A, the optimum design of Chart 1A should be used.

For this B_m-limited method, consider the simultaneous solution of equations 1, 2, 3 and

$$B_m = (\text{constant}) \tag{23}$$

The solution is

$$\beta = \mu_{\Delta g}[g_1{}^2 - (1/\mu_{\Delta l}')] \tag{24}$$

and

$$\sqrt[4]{n_1} = \frac{B_m\sqrt{g_1}}{\mu_l} + \left(\frac{H_cB_m}{B_r} + H_c\right)\left(g_1{}^2 - \frac{1}{\mu_{\Delta l}'}\right)$$

$$\mu_{\Delta g}\sqrt{g_1} \tag{25}$$

For a constant B_m, this equation can be tabulated in a reverse manner similar to equation 4 or equation 21. First a value of B_m is assumed for the particular chart under construction. Then for various B_Δ/B_m ratios, various values of g_1 are assumed and the resulting n_1 values calculated by equation 25. Equation 24 can be used to calculate β for every point. For each point, the I_Δ/I_m ratio can be found by equations 2 and 3 and curves of g_1 and β versus n_1 can be drawn for the various I_Δ/I_m ratios. This is Chart 1C. The n_{1r}, g_{1r}, and s axes are identical to those on Charts 1A and 1B. Fig. 3 is Chart 1C for a B_m value of 5 kilogausses.

If the B_m from Chart 1A for a given set of specifications is too large, the designer would use Chart 1C. Equation 11 still applies for calculating ℓ_t and equation 12 still applies to N.

The Design Method for Inductors Limited by Copper Loss Heating-Power Density

MINIMUM SIZE DESIGN

In some cases the circuit requirements limiting the resistance are so loose that the only factor limiting the resistance of the inductor is the heating of the inductor by the current through its copper resistance. For these cases it is convenient to prepare charts for the design of inductors subject only to the limitation of their own heating.

The average rate of heat generation in the copper resistance is $I^2_{\text{rms}}R_c$. The deterioration or destruction of the inductor is a function of the temperature which, in turn, is a function of the density of thermal power passing through the surface of the unit. For units of the same shape, the surface area may be expressed by dimensionless constant, p.

When N has been eliminated from the inductance and thermal power density equations

$$K_c = \frac{u\rho LI^2_{\text{rms}}[(1/\mu_{\Delta 1}) + (\beta/\mu_{\Delta g})]10^8}{0.4\pi pvw\ell_t{}^4} \tag{26}$$

This equation is similar to equation 1 in the location of the various factors in numerator and denominator. The optimum design is again that for which $[(1/\mu_{\Delta l}') + (\beta/\mu_{\Delta g})]$ is minimum. The system of equations 2, 3, and 26 may be solved for minimum $[(1/\mu'_{\Delta i})+(\beta/\mu_{\Delta g})]$ in a manner analogous to that of the resistance-limited method. The solution is

$$\sqrt[8]{n_2}=\frac{(B_m+B_r)\left[B_mD_n-\mu_{\Delta g}\left(H_c+\frac{H_cB_m}{B_r}\right)D_\Delta\right]-\frac{B_r}{\mu_i}}{\sqrt[8]{\frac{B_r}{H_c\mu_{\Delta g}}\left[B_mD_n-\mu_{\Delta g}\left(H_c+\frac{H_cB_m}{B_r}\right)D_\Delta-\frac{1}{\mu_i}\right]+\frac{1}{\mu_{\Delta i}}}} \qquad (27)$$

where

$$n_2=(0.4\pi)^7p^3v^3LI_m^8K_c10^8/[u^3w\rho^3I_{rms}^6] \qquad (28)$$

Equation 27 is analogous to equation 4 for setting up a design chart. For various B_m values, n_2 can be calculated; and a plot of B_m versus n_2 can be made in a manner similar to that of Chart 1A.

A curve for β can be also added to the chart by use of equation 8.

The magnetic length

$$\ell_i=\sqrt[4]{[(1/\mu_{\Delta i}')+(\beta/\mu_{\Delta g})]u\rho LI_{rms}^2 10^8/[0.4\pi pvwK_c]} \qquad (29)$$

so that it is convenient to plot an application function g_2 defined by

$$g_2\triangleq\sqrt[4]{(1/\mu_{\Delta i}')+(\beta/\mu_{\Delta g})} \qquad (30)$$

With this definition,

$$\ell_i=g_2\sqrt[4]{u\rho10^8/[0.4\pi pvw]}\,\sqrt[4]{LI_{rms}^2/K_c} \qquad (31)$$

The value of g_2 may be calculated by inserting each β value of equation 8 into equation 30. These values can be added to the chart, and equation 31 can be used to compute ℓ_i. The formula for N is

$$N=g_2^{3/2}\sqrt[8]{\frac{pvL^3K_c10^{24}}{(0.4\pi)^3uw^3\rho I_{rms}^2}} \qquad (32)$$

The copper resistance may be calculated as

$$R_c=g_2^2\sqrt{\frac{pu\rho10^8}{0.4\pi vw}}\sqrt{\frac{LK_c}{I_{rms}^2}}. \qquad (33)$$

This equation may be used for calculating the winding resistance of a power-density-limited inductor without completing the other design calculations.

The design chart for power-density-limited inductors can be set up in terms of the I_Δ/I_m ratios in a manner similar to Chart 1A. The result is shown in Chart 2A, Fig. 4.

Auxiliary Axes and Variable Stack Heights

For designing with particular values of u, v, w, and ρ, it is convenient to set up additional axes similar to the n_{1r} and g_{1r} axes used in the resistance-limited design charts. First consider

$$n_{2r}\triangleq LI_m^8K_c^3/I_{rms}^6=\{u^3w\rho^310^{-8}/[(0.4\pi)^7p^3v^3]\}n_2 \qquad (34)$$

A logarithmic n_{2r} axis can be laid out parallel to the logarithmic n_2 axis to perform the multiplication indicated in equation 34. The designer calculates n_{2r} by equation 34 and enters the chart at the corresponding point of the n_{2r} axis. The calculation of n_{2r} by equation 34 is simpler than the calculation of n_2 by equation 28; but the n_2 axis on the chart is a function only of the magnetic properties of the material, in equation 27, while the n_{2r} axis is dependent upon u, v, ρ, w, and p.

For simplifying the ℓ_i calculation, another application function g_{2r} is defined.

$$g_{2r}\triangleq\ell_i\sqrt[4]{K_c/(LI_{rms}^2)}=g_2\sqrt[4]{(u\rho10^8)/(0.4pvw)} \qquad (35)$$

This simplifies the calculation of equation 31, and a logarithmic g_{2r} axis can be laid out parallel to the logarithmic g_2 axis to do this operation directly.

The constants in equations 34 and 35 may be evaluated for the scrapless *EI* low-voltage case, as were the constants in the resistance-limited design. For military standard (*MIL-T-27*) cases,

$$p=1.3 \qquad (36)$$

From equation 34

$$n_{2r}/n_2=u^3w\rho^310^{-8}/[(0.4\pi)^7p^3v^3]=5\times10^{-22} \qquad (37)$$

and from equation 35

$$g_{2r}/g_2=\sqrt[4]{u\rho10^8/[0.4\pi pvw]}=30 \qquad (38)$$

As in the resistance-limited charts, an s axis may be set up to facilitate the design

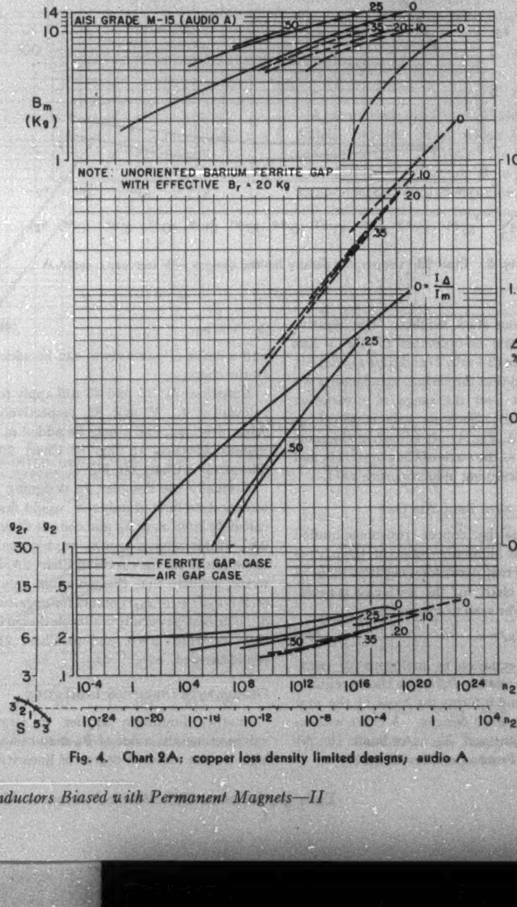


Fig. 4. Chart 2A: copper loss density limited designs; audio A

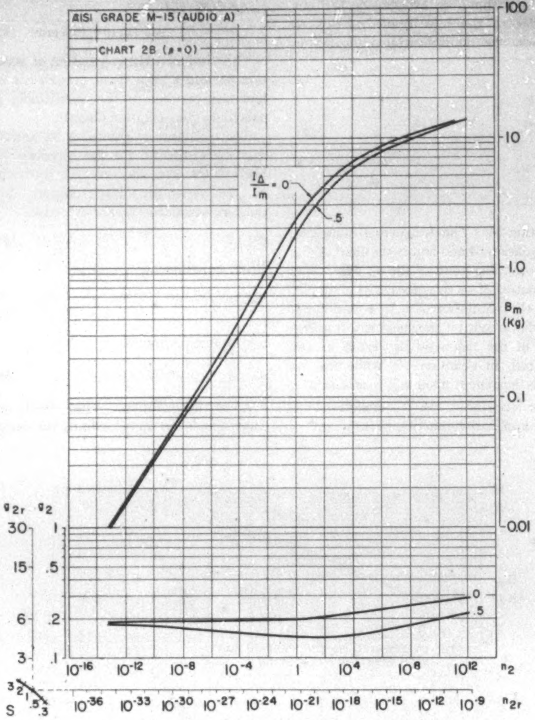


Fig. 5. Chart 2B: copper loss density limited designs with zero gap; audio A

for various stack heights. The negative logarithmic co-ordinates of the s axis points are $(2+s)^3s/27$ and $[(2+s)/(3s)]^{1/4}$, provided that the surface area factor p is a constant over the range of s variation considered. Such an axis is shown in Chart 2A, Fig. 4. It is used in the same manner as the s axis of Charts 1A, 1B, and 1C for designing inductors with $s \neq 1$.

Design with Zero Air Gap

As in the copper resistance-limited case, the minimum value of $[(1/\mu_{\Delta i}')+(\beta/\mu_{\Delta g})]$ may occur for a negative value of β, so a chart for the $\beta=0$ case is necessary. The system solution is

$$B_m{}^8\mu_{\Delta i}'/\mu_i{}^8=n_2 \tag{39}$$

This equation is analogous to equation 4 in relating B_m to the known requirements, and it forms the basis of the new chart for $\beta=0$ designs. A chart with n_2 as abscissa and B_m as ordinate can be made. From equations 20 and 30,

$$g_2=\sqrt[4]{1/\mu_{\Delta i}'} \tag{40}$$

and a curve of g_2 versus n_2 can be added to the chart.

Equations 31, 32, and 33 still apply for calculating l_i, N, and R_c, respectively. Axes of n_{2r}, g_{2r}, and s may be added in a manner identical to that of Chart 2A. The result is Chart 2B, Fig. 5.

Normally the designer, in designing a temperature-limited inductor, would first calculate index n_2 or n_{2r} and consult Chart 2A. If this value of n_2 or n_{2r} was too small to be found on the curves of Chart 2A, it would indicate that the minimum value of $[(1/\mu_{\Delta i}')+(\beta/\mu_{\Delta g})]$ occurs for negative β. The best physically realizable design is that for which $\beta=0$, and so Chart 2B would be used.

Design with Induction Limitation

Another limitation applies to large values of n_2; the value of B_m may be too high for the required degree of linearity, or $[(1/\mu_{\Delta i}')+(\beta/\mu_{\Delta g})]$ may not have a minimum. For such cases, charts for the optimum designs with fixed B_m values are useful.

The solution of the equations 2, 3, and 26 with $\beta=0$ is

$$\beta=\mu_{\Delta g}[g_2{}^4-(1/\mu_{\Delta i}')] \tag{41}$$

and

$$\sqrt[8]{n_2}=\frac{B_m\left[\frac{1}{\mu_i}+\mu_{\Delta g}\left(\frac{H_c}{B_r}+\frac{H_c}{B_m}\right)\left(g_2{}^4-\frac{1}{\mu_{\Delta i}'}\right)\right]}{\sqrt{g_2}} \tag{42}$$

This equation is tabulated by assuming a constant B_m value for the particular chart under construction. Then, for various B_Δ/B_m ratios, various values of g_2 are assumed, and the resulting n_2 values calculated by equation 42. Curves with n_2 as abscissas and g_2 as ordinates can then be plotted with I_Δ/I_m as a constant by the method used for previous charts. By means of equation 41, curves of β versus n_2 can be constructed. Such a chart is called Chart 2C. This also has the n_{2r}, g_{2r}, and s axes identical to those of Charts 2A and 2B. Fig. 6 is Chart 2C for a B_m value of 5 kilogausses. Chart 2C would normally be used after the designer has used Chart 2A and found the optimum B_m value to be too high. Equations 31, 32, and 33 would be used to calculate l_i, N, and R_c, respectively.

The Design Method for Inductors Limited by Total Resistance

General Equations

The first synthesis section considered the problem of synthesis of inductors where only the copper resistance was specified. In any iron core inductor, there is core loss when the induction or magnetizing force is varying with time. This core loss is generally a function of the iron, the induction wave (including frequency, waveform, and amplitude), and the volume of the iron having this waveform. For the type of core materials considered here, the core loss is expressed approximately by

$$W_i=\gamma B_\Delta{}^2wl_i{}^3 \tag{43}$$

where γ is a constant for a given material, frequency, and waveform.

The loss is supplied by the time-varying components of current and voltage, so that in computing the resistance which this loss presents in the circuit, the mean square current used must not include the d-c component.

The effective iron loss resistance is

$$R_i=0.4\pi\gamma I_\Delta{}^2L10^8/\\ \{(I_{\text{rms}}{}^2-I^2)[(1/\mu_{\Delta i})+(\beta/\mu_{\Delta g})]\} \tag{44}$$

The total resistance to the time-varying component of current is

$$R_{ac}=\frac{u\rho L10^8\left(\dfrac{1}{\mu_{\Delta i}}+\dfrac{\beta}{\mu_{\Delta g}}\right)}{0.4\pi vw\ell_i^2}+\frac{0.4\pi\gamma I_\Delta^2L10^8}{(I_{rms}^2-I^2)\left(\dfrac{1}{\mu_{\Delta i}}+\dfrac{\beta}{\mu_{\Delta g}}\right)}\quad(45)$$

The equation has a minimum when the two terms, copper resistance and iron loss resistance, are equal.

To have equal resistances

$$(1/\mu_{\Delta i})+(\beta/\mu_{\Delta g})=\sqrt{(0.4\pi)^2vw\gamma\ell_i^2I_\Delta^2/[u\rho(I_{rms}^2-I^2)]}\quad(46)$$

It has previously been shown that $[(1/\mu_{\Delta i})+(\beta/\mu_{\Delta g})]$ has a minimum value which is a function of the magnetic core material and the electrical specifications, L, I_m, I_Δ, and R_c. This minimum value of $[(1/\mu_{\Delta i})+(\beta/\mu_{\Delta g})]$ may be less than, equal to, or greater than the value given by equation 46 for minimum R_{ac}. If the minimum value of $[(1/\mu_{\Delta i})+(\beta/\mu_{\Delta g})]$ is greater than the value given by equation 46, the minimum R_{ac} condition cannot be realized, and from equations 44 and 45 it is apparent that

$$R_i<(R_{ac}/2)<R_c<R_{ac}\quad(47)$$

More than half the effective resistance is due to the winding loss. On the other hand, if the minimum value of $[(1/\mu_{\Delta i})+(\beta/\mu_{\Delta g})]$ is less than, or equal to, the value given by equation 46, the minimum R_{ac} condition can be realized. The components of resistance due to the winding loss and core loss are equal. In no case where R_{ac} is a controlling factor is it advantageous to design for a core loss component of resistance greater than the winding loss component of resistance.

Consider the case in which the optimum value of $[(1/\mu_{\Delta i})+(\beta/\mu_{\Delta g})]$ can be realized.

Then

$$\ell_i=[2L10^8/R_{ac}]\sqrt{[u\rho\gamma I_\Delta^2]/[vw(I_{rms}^2-I^2)]}\quad(48)$$

Inserting this value of ℓ_i into equation 46 gives

$$(1/\mu_{\Delta i})+(\beta/\mu_{\Delta g})=0.4\pi\gamma I_\Delta^2L10^8/[(I_{rms}^2-I^2)(R_{ac}/2)]\quad(49)$$

This equation may be used with Chart 1A, or with Charts 1B or 1C where their respective conditions apply, to determine in which case a given set of specifications lie. The value of index n_1 (or n_{1r}) may be calculated from the known specifications and an assumed value of winding resistance.

$$R_c=(R_{ac}/2)\quad(50)$$

From the chart the value of g_1 is read from the appropriate curve.

Equation 49 can be used to calculate the value of $[(1/\mu_{\Delta i})+(\beta/\mu_{\Delta g})]$ for minimum R_{ac}. This value will be symbolized by G_1^2 so that

$$G_1^2=\frac{0.4\pi\gamma I_\Delta^2L10^8}{(I_{rms}^2-I^2)(R_{ac}/2)}\quad(51)$$

A comparison of g_1^2 and G_1^2 will immediately indicate under which case the design will fall. If $G_1^2<g_1^2$ the value of $[(1/\mu_{\Delta i})+(\beta/\mu_{\Delta g})]$ for minimum R_{ac} cannot be reached and the optimum design will be one of the copper-resistance-

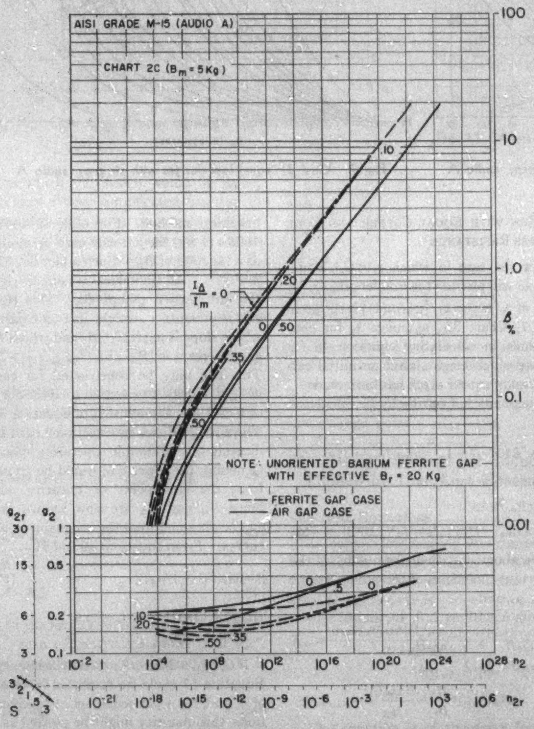


Fig. 6. Chart 2C: copper loss density limited designs with five kilogauss maximum induction, audio A

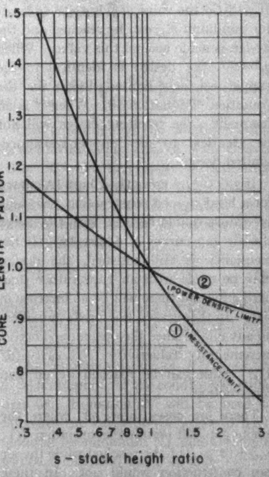


Fig. 7. Chart 3A: core length factors for scrapless EI laminations

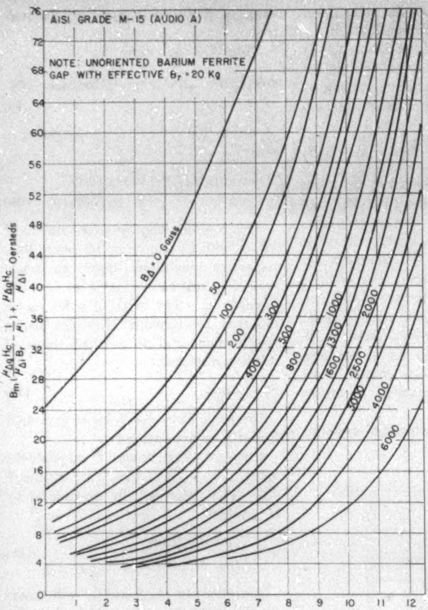


Fig. 8. Chart 3: equal loss designs with ferrite gap; audio A

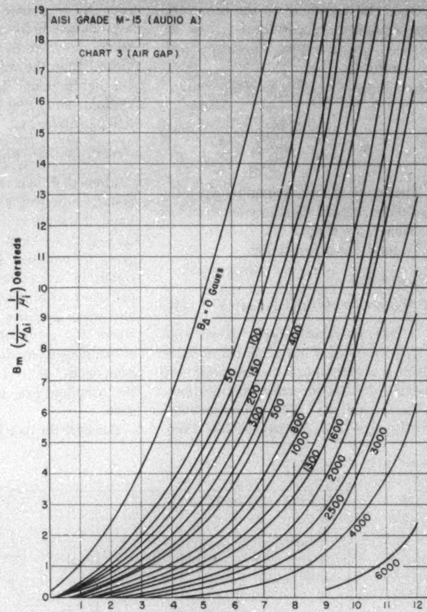


Fig. 9. Chart 3: equal loss designs with air gap; audio A

Vol.
Pt. 1
196

limited designs given by Chart 1A, 1B, or 1C, the specific details to be given later. If $g_1^2 \leqq G_1^2$ the value of $[(1/\mu_{\Delta i})+(\beta/\mu_{\Delta g})]$ for minimum R_{ac} can be reached and the design is made around this value. When $g_1^2=G_1^2$, the design is identical to the copper-resistance-limited design for the condition of equation 50. When it is an inequality, the resulting design will not be identical to any copper-resistance-limited design.

If $g_1^2<G_1^2$ by using the Chart 1A curves for a hard-gapped construction, the same procedure should be tried with the use of the air gap curves of Chart 1A. If the inequality is still satisfied, the design may be made with air gaps rather than hard gaps and there will be no increase in size. In these cases, the ferrite gap offers no advantage since the optimum effective incremental reluctivity $[(1/\mu_{\Delta i})+(\beta/\mu_{\Delta g})]$ value can be reached by air gap construction.

There are cases, however, where the ferrite gap will permit the optimum $[(1/\mu_{\Delta i})+(\beta/\mu_{\Delta g})]$ to be reached while the air gap construction would not. In these cases the ferrite-gapped construction will be smaller.

Design with Equal Copper and Core Loss Resistance

For the case in which $g_1^2<G_1^2$, equation 48 will provide the iron length, ℓ_i. A plot of $\sqrt{u(s)w(1)/[w(s)u(1)]}$ is shown in Fig. 7, Chart 3A, as curve 1, for convenience in calculating equation 48 for various stack heights after an initial calculation of square stack has been made.

The number of turns may be calculated from

$$N=\sqrt{L10^8}\ \sqrt[4]{v\gamma I_\Delta^2/[uw\rho(I_{\mathrm{rms}}^2-I^2)]} \qquad (52)$$

The equation for B_Δ is

$$B_\Delta=(R_{ac}/2)^2\times \sqrt[4]{v^3w(I_{\mathrm{rms}}^2-I^2)^5/[u^3\rho^3\gamma^5 I_\Delta^6(L10^8)^6]} \qquad (53)$$

The above equations formulate all the important quantities of the design except β, B_m, μ_i, and $\mu_{\Delta i}$.

From equations 2, 3, 49, and 51

$$B_m\left(\frac{\mu_{\Delta g}H_c}{\mu_{\Delta i}B_r}-\frac{1}{\mu_i}\right)+\frac{\mu_{\Delta g}H_c}{\mu_{\Delta i}} = B_m\mu_{\Delta g}G_1^2\frac{H_c}{B_r}-\frac{B_\Delta I_m G_1^2}{I_\Delta}+\mu_{\Delta g}G_1^2H_c \qquad (54)$$

This equation may be solved by a graphical method. For each of several values of B_Δ, the left side may be plotted as a function of B_m. Such a plot is Chart 3, Fig. 8. All the quantities on the right side are known except B_m. This right side represents a straight line on Chart 3 whose slope is $\mu_{\Delta g}H_cG_1^2/B_r$ and whose B_m axis intercept is $[B_\Delta I_m B_r/(I_\Delta \mu_{\Delta g} H_c)]-B_r$. This line may be constructed for each design, and its intersection with the curve of Chart 3 is the solution of equation 54. There may be two intersections, each indicating an allowable operating point. Generally the lower one would be chosen from the standpoint of linearity, etc. Since B_m and B_Δ are now known, $\mu_{\Delta i}$ can be obtained from the original iron $\mu_{\Delta i}$ curves. From equations 49 and 51

$$\beta=\mu_{\Delta g}[G_1^2-(1/\mu_{\Delta i})] \qquad (55)$$

Design Limited by Copper Resistance

If $G_1<g_1$, a different attack is necessary. Equation 51 could be rewritten in terms of n_1 and, for a particular set of specifications, this quantity might be plotted as a curve on Chart 1A. Its intersection with the g_1 curve would be the design point

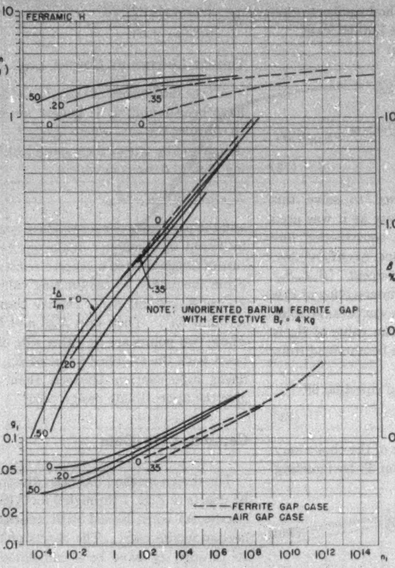


Fig. 10. Chart 1A: copper resistance limited designs, Ferramic H

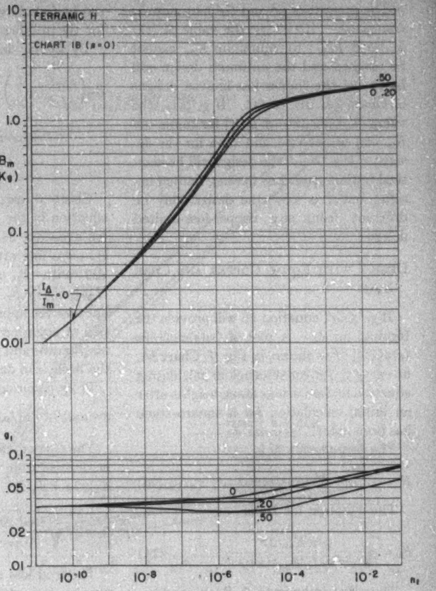


Fig. 11. Chart 1B: copper resistance limited designs with zero gap, Ferramic H

from which the design would be made, but it is a very complicated function to compute, and an approximate method will be satisfactory in most cases. From inequality 47 the limiting values that R_c can have are $R_{ac}/2$ and R_{ac}. Calculations can be made for these two points as follows:

Let

$$R_c'=(R_{ac}/2) \tag{56}$$

Calculate index n_1' corresponding to R_c' by equation 5. Read g_1' corresponding to n_1'. Calculate, from equation 44 and 10,

$$R_i'=0.4\pi\gamma I_\Delta^2 L10^8/[(I_{\mathrm{rms}}^2-I^2)(g_1')^2] \tag{57}$$

and

$$R_{ac}'=R_c'+R_i' \tag{58}$$

Let

$$R_c''=R_{ac} \tag{59}$$

In the same manner as before calculate n_1'', read g_1'', and calculate R_i'' and R_{ac}''. By consideration of equation 45, it is seen that

$$R_{ac}'<R_{ac} \tag{60}$$

As long as

$$I_\Delta\neq 0 \tag{61}$$

and

$$\gamma\neq 0 \tag{62}$$

$$R_i''>0 \tag{63}$$

so that

$$R_{ac}''>R_{ac} \tag{64}$$

These are then two points on either side of the point where the inductor design will have the proper R_{ac}. Linear interpolations may be used to find the approximate design point.

The Design Method for Inductors Limited by Total Loss Heating

General Equations

As copper loss heating was considered in an earlier section, heating by the total resistance will be considered now. The total power density is

$$K_{ac}=\frac{u\rho I_{\mathrm{rms}}^2 L10^8[(1/\mu_{\Delta i})+(\beta/\mu_{\Delta g})]}{0.4\pi p v w \ell_i^4}+\frac{0.4\pi I_\Delta^2 L10^8}{p\ell_i^2[(1/\mu_{\Delta i})+(\beta/\mu_{\Delta g})]} \tag{65}$$

The equation is of the same form as equation 45 and will be treated in the same way.

For equal loss case

$$\ell_i=\sqrt[3]{\frac{I_{\mathrm{rms}}I_\Delta L10^8}{p(K_{ac}/2)}\sqrt{\frac{u\rho\gamma}{vw}}} \tag{66}$$

and

$$\frac{1}{\mu_{\Delta i}}+\frac{\beta}{\mu_{\Delta g}}=0.4\pi\sqrt[3]{\frac{vw\gamma I_\Delta^4 L10^8}{pu\rho I_{\mathrm{rms}}^2(K_{ac}/2)}} \tag{67}$$

This equation can be used for comparison with Chart 2A, or with Charts 2B or 2C when they apply, to determine whether this value $[(1/\mu_{\Delta i})+(\beta/\mu_{\Delta g})]$ can be realized. From equation 28 a value of index n_2 may be calculated for the condition that

$$K_c=(K_{ac}/2) \tag{68}$$

where K_{ac} is determined by the allowable temperature rise. The chart will then give application function g_2.

The minimum value of $[(1/\mu_{\Delta i})+(\beta/\mu_{\Delta g})]$ is obtained for the condition of equation 68. Equation 67 will be used to calculate the value of $[(1/\mu_{\Delta i})+(\beta/\mu_{\Delta g})]$ required for minimum K_{ac} and this value will be symbolized by G_2^4 so that

$$G_2^4=0.4\pi\sqrt[3]{\frac{vw\gamma^2 I_\Delta^4 L10^8}{pu\rho I_{\mathrm{rms}}^2(K_{ac}/2)}} \tag{69}$$

The quantities g_2^4 and G_2^4 may now be compared. If $G_2^4 < g_2^4$ the value of $[(1/\mu_{\Delta t})+(\beta/\mu_{\Delta g})]$ for minimum K_{ac} cannot be reached and the optimum design will be one of the copper-loss-limited designs of Chart 2A, 2B, or 2C. If $g_2^4 \leqq G_2^4$ the value of $[(1/\mu_{\Delta t})+(\beta/\mu_{\Delta g})]$ for minimum K_{ac} can be realized and used for the design. If $g_2^4 = G_2^4$, the design will be identical to the copper loss heating limited design; but if $g_2^4 < G_2^4$, the design may be different from any copper-loss-limited design.

Design with Equal Copper and Core Loss

If $g_2^4 \leqq G_2^4$, equation 66 will provide the formula for ℓ_t. A plot of $\{u(s)w(1)/[w(s)u(1)]\}^{1/6}$ is shown in Fig. 7, Chart 3A, as curve 2, for convenience in calculating equation 66 for various stack heights after an initial calculation for a square stack has been made.

The formula for N is

$$N=\sqrt{L10^8}\ \sqrt[4]{v\gamma I_\Delta^2/[uw\rho I_{rms}^2]} \qquad (70)$$

The equation for B_Δ is

$$B_\Delta=\sqrt[12]{\frac{p^8 v(K_{ac}/2)^8}{uw^6\rho\gamma^7 I_\Delta^2 I_{rms}^2 (L10^8)^2}} \qquad (71)$$

The other unknowns, β, B_m, μ_t, and $\mu_{\Delta t}$ can be found by a method similar to that for the total-resistance-limited case.

By similar derivation

$$B_m\left(\frac{\mu_{\Delta g}H_c}{\mu_{\Delta t}B_r}-\frac{1}{\mu_t}\right)+\frac{\mu_{\Delta g}H_c}{\mu_{\Delta t}} = B_m\mu_{\Delta g}G_2^4\frac{H_c}{B_r}-\frac{B_\Delta I_m G_2^4}{I_\Delta}+\mu_{\Delta g}G_2^4 H_c \qquad (72)$$

Chart 3 can be used to solve this equation in the same way as it was used for equation 54. The construction line is now a line with a slope, $(\mu_{\Delta g}H_c G_2^4)/B_r$, and with B_m axis intercept $[B_\Delta I_m B_r/(I_\Delta \mu_{\Delta g} H_c)]-B_r$. The intersection of this line and the appropriate curve is the solution of equation 72. The value of B_m is now known and $\mu_{\Delta t}$ can be obtained from the basic iron data.

From equation 67 and equation 69

$$\beta=\mu_{\Delta g}[G_2^4-(1/\mu_{\Delta t})] \qquad (73)$$

The copper resistance of the choke may be derived from equation 33 by inserting $K_{ac}/2$ for K_c.

$$R_c=G_2^2\sqrt{\frac{pu\rho L10^8(K_{ac}/2)}{0.4\pi vwI_{rms}^2}} \qquad (74)$$

The iron loss component of resistance may be derived from equation 44 as

$$R_t=\frac{0.4\pi\gamma I_\Delta^2 L10^8}{G_2^4(I_{rms}^2-I^2)} \qquad (75)$$

Design Limited by Copper Loss

All the dimensions of the $g_2^4 \leqq G_2^4$ design have been found, so it is next necessary to consider the case where $G_2^4 < g_2^4$. This design will be one of the copper loss heating limited designs in which

$$K_t<(K_{ac}/2)<K_c<K_{ac} \qquad (76)$$

and the particular one for smallest size will be found in a manner analogous to the case of design limited by copper resistance. From inequality 76 the limiting values of K_c are $K_{ac}/2$ and K_{ac}. Denote these by K_c' and K_c''. Then find K_{ac}' and K_{ac}''. These two points are on either side of the point with the specified value of K_{ac} and linear interpolation between them will give the approximate design point.

Ceramic Core Data

The design method has been illustrated by charts of Figs. 1 through 9 for AISI-15 (Audio A) cores biased with unoriented barium ferrite gaps. The charts of Figs. 10 through 17 have been calculated for Ferramic H cores. These charts are similar to those for the metallic cores ex-

Vol.
Pt. 1
196

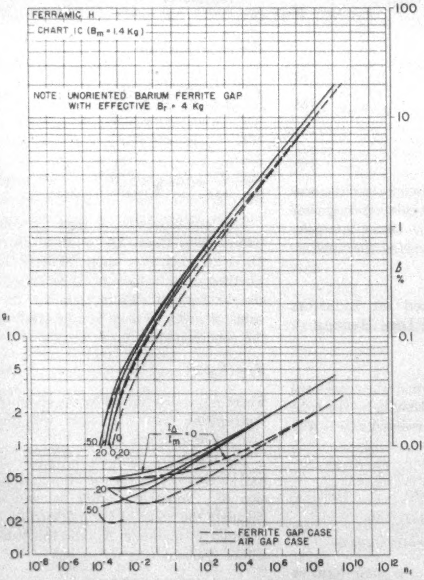


Fig. 12. Chart 1C: copper resistance limited designs with 1.4 kilogauss maximum induction, Ferramic H

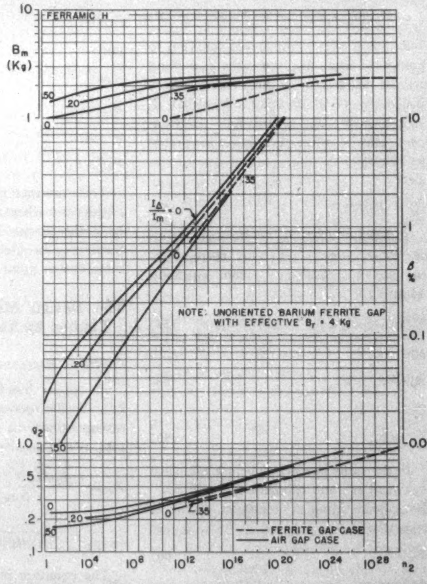


Fig. 13. Chart 2A: copper loss density limited designs, Ferramic H

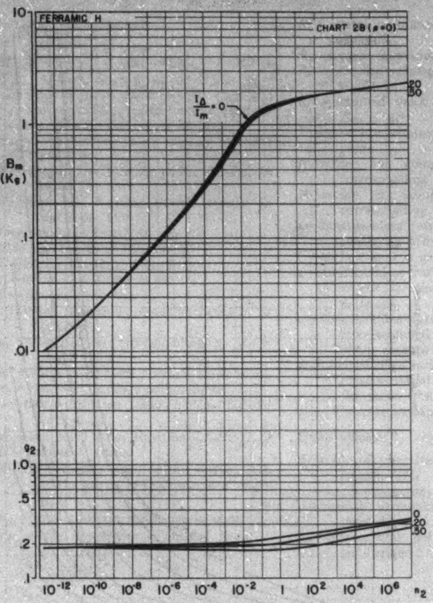


Fig. 14. Chart 2B: copper loss density limited designs with zero gap, Ferramic H

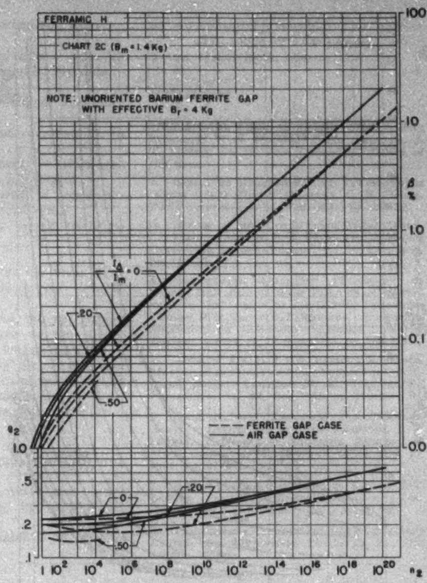


Fig. 15. Chart 2C: copper loss density limited designs with 1.4 kilogauss maximum induction, Ferramic H

cept that the gap angling was such that the gap inductions were one half those of the main core, and the auxiliary axes are omitted because these ceramic cores are neither standardized nor laminated.

General Design Method

A general map of the synthesis method is shown in Fig. 18. Where the circuit imposes a resistance limitation, the design method follows the course in the upper part of the map. In the general case, the total resistance R_{ac} would be given and the method of design would be chosen by a test inequality. In some cases, only the copper d-c resistance R_c is specified and the R_c limited design method would be used.

Within the R_c-limited area, there are several charts. Normally Chart 1A is used first since it specifies the minimum size design otherwise unrestricted. It has a lower limit below which the air gap would be negative; and in this range, Chart IB will specify the zero air gap design, which is the best physically realizable design. Chart 1A has upper limits where a minimum ceases to exist, and in many cases considerations of linearity may restrict the maximum induction to be used. For design in these ranges, the Charts 1C provide designs.

Within the $R_c=R_t$ design area, the design method involves more algebraic equations and fewer curves than the R_c-limited area.

When the design has been made to meet a specified resistance figure, the thermal power density, which determines the temperature rise, should be checked. If the thermal power density of the design is too high or if there is no circuit requirement on the allowable resistance, the design should be made in the thermal power density limited area shown in the lower portion of Fig. 18.

The map of this area is similar to the map for specified resistance. Normally the designer would determine the specific method by means of a test inequality, although in the case of a small alternating current and large direct current, he would use the K_c-limited design immediately. There are several charts within the K_c-limited area, and these are analogous to the charts in the R_c-limited area. The design method within the $K_c=K_t$ area is similar to the method in the $R_c=R_t$ area and uses the same Chart 3. In most cases, the design synthesized by the method will be subject to a winding calculation, taking into account the actual wire size, local winding methods, etc.

The general method, insofar as the basic ideas are concerned, may also be used for transformers with superposed a-c and d-c magnetization. When so used, all the currents should be referred to one winding, and the design can then be based on the magnetizing current and the open circuit inductance requirements. The heating effects must be based on the total magnetizing and load currents. The copper area factor v must be that for the reference winding alone. In a 2-winding transformer, the value of v would be approximately half of the value used in the inductor design for the same winding methods since each coil should occupy half the window.

Conclusions

The theory developed indicates the feasibility of the ferrite-gapped inductor for miniaturization of inductive devices carrying unsymmetrical currents.

The mathematical relations developed

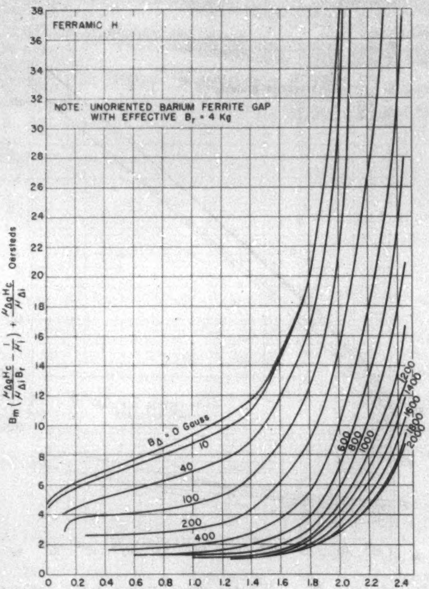


Fig. 16. Chart 3: equal loss designs with ferrite gap; Ferramic H

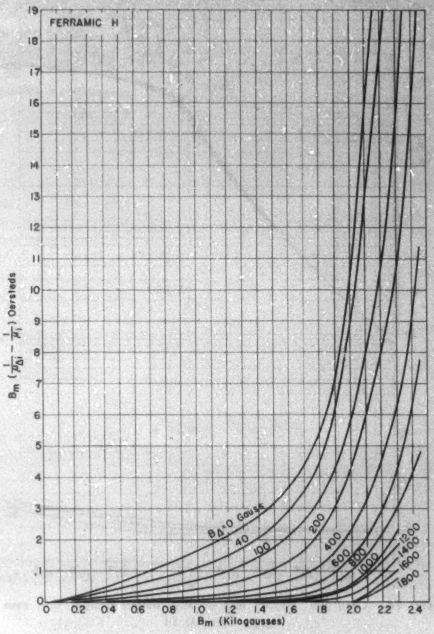


Fig. 17. Chart 3: equal loss designs with air gap; Ferramic H

provide a method of constructing design charts directly from the magnetic material characteristics. Those charts may then be used for the design (synthesis) of biased inductors without the usual trial-and-error procedures for the first design.

The charts provide a direct evaluation of the applicability of the various ferromagnetic materials. If, for example, for a given value of index n_1, two materials have application function numbers, g_1, which differ by a factor of two, the resulting optimum inductors constructed from these materials would have a factor of two ratio between their linear dimensions and a ratio of eight (2^3) between their weights and their volumes.

A secondary use of the method is that of providing a quick size estimate for the equipment designer who may need to weigh inductor specifications and size against the specifications and size of other components with which the inductor is connected.

The method is limited by the quantity and quality of the magnetic data available. No complete set of incremental data has been available from any one steel manufacturer in recent years, nor is the range of variation of these data generally available.

Nomenclature

Subscripts on Time Variable Quantities (period P)

m = maximum value during a cycle

rms = root-mean-square, $Z_{rms} = \sqrt{(1/P)\int_0^P z^2 dt}$

Δ = incremental, $Z_\Delta = (Z_{max} - Z_{min})/2$
no subscript = time average, $Z = (1/P)\int_0^P z\,dt$

General Subscripts

c = value within the coil (copper)
g = value within the gap
i = value within the iron

Letter Symbols

a_c = net copper cross-sectional area in cm²
a_i = net iron cross-section area in cm²
b = time variable induction in gausses
B_r = gap material effective residual induction in gausses
$D_\Delta = (1/\mu_\Delta'^2)d\mu_\Delta'/dB_m$
$D_N = (1/\mu^2)d\mu/dB_m$
g_1 = application function $\sqrt{(1/\mu_\Delta)+(\beta/\mu_{\Delta g})}$
g_1' = value of g_1 corresponding to $R_c' = R_{ac}/2$
g_{1r} = simplified application function, $l_i\sqrt{R_c/L}$
g_1'' = value of g_1 corresponding to $R_c'' = R_{ac}$
G_1 = value of g_1 for which $R_c = R_i$

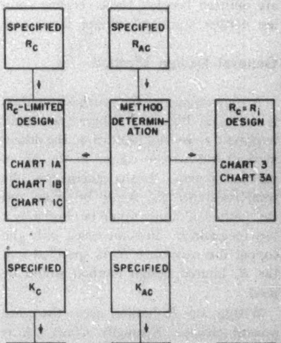

SPECIFIED R_C
SPECIFIED R_{AC}
R_C-LIMITED DESIGN
CHART 1A
CHART 1B
CHART 1C
METHOD DETERMINATION
$R_C = R_i$ DESIGN
CHART 3
CHART 3A
SPECIFIED K_C
SPECIFIED K_{AC}
K_C-LIMITED DESIGN
CHART 2A
CHART 2B
CHART 2C
METHOD DETERMINATION
$K_C = K_i$ DESIGN
CHART 3
CHART 3A

Fig. 18. Map of synthesis method

Vol.
Pt. 1
196

g_2 = application function $\sqrt[4]{(1/\mu_\Delta')+(\beta/\mu_{\Delta g})}$

g_{2r} = simplified application function $\sqrt[4]{K_c/(LI_{rms}^2)}$

G_2 = value of g_2 for which $K_c = K_t$

H_c = gap material coercive force in oersteds

i = time variable current in amperes

K_{ac} = total heating power density in watts per cm^2

K_c = copper loss heating power density in watts per cm^2

K_t = iron loss heating power density in watts per cm^2

ℓ = generalized length in cm

ℓ_c = mean length of turn in cm

ℓ_g = effective length of gap in cm

ℓ_t = mean iron length in cm

L = inductance in henrys

n_1 = index function $(0.4\pi)^8 v^2 w I_m^4 R_c^2 10^{-8}/(u^2\rho^2 L)$

n_1' = value of n_1 corresponding to $R_c' = R_{ac}/2$

n_1'' = value of n_1 corresponding to $R_c'' = R_{ac}$

n_{1r} = simplified index function $I_m^4 R_c^2/L$

n_2 = index function $(0.4\pi)^7 p^2 v^3 L I_m^8 K_c^2 10^9/(u^2 w \rho^3 I_{rms}^6)$

n_{2r} = simplified index function $LI_m^8 K_c/I_{rms}^6$

N = number of turns

p = dimensionless ratio of the surface area to the square of the mean iron length S/ℓ_t^2

R_{ac} = total effective resistance in ohms

R_{ac}' = total resistance calculated from $R_c' = R_{ac}/2$

R_{ac}'' = total resistance calculated from $R_c'' = R_{ac}$

R_c = copper resistance in ohms

R_c' = assumed value of copper resistance $R_{ac}/2$

R_c'' = assumed value of copper resistance R_{ac}

R_t = core loss component of resistance in ohms

R_t' = core loss component of resistance corresponding to $R_c' = R_{ac}/2$

R_t'' = core loss component of resistance corresponding to $R_c'' = R_{ac}$

s = dimensionless ratio of stack height to center leg width

S = surface area in cm^2

u = dimensionless ratio of the mean turn length to the mean iron length, ℓ_c/ℓ_t

v = dimensionless ratio of the net copper cross-sectional area to the square of the mean iron length, a_c/ℓ_t^2

w = dimensionless ratio of the net iron core cross-sectional area to the square of the mean iron length, a_c/ℓ_t^2

W_t = iron core loss in watts

Greek Symbols

β = ratio of effective gap length to iron length, ℓ_g/ℓ_t

γ = core loss constant in watts/(cm^3 gauss^2)

μ = normal permeability in gausses per oersted

μ_Δ = incremental permeability in gausses per oersted

μ_Δ' = incremental permeability as a function of B_m alone for a fixed B_Δ/B_m ratio

ρ = coil resistivity in ohm-cm

References

1. INDUCTORS BIASED WITH PERMANENT MAGNETS: PART I—THEORY AND ANALYSIS, J. T. Ludwig. *AIEE Transactions*, pp. 273–78 of this issue.

2. DESIGN OF REACTANCES AND TRANSFORMERS WHICH CARRY DIRECT CURRENT, C. R. Hanna. *AIEE Journal*, vol. 46, Feb. 1927, pp. 128–31.

3. DESIGN OF IRON CORED INDUCTANCES CARRYING DIRECT CURRENT, N. H. Crowhurst. *Electronic Engineering*, London, England, vol. 22, Dec. 1950, pp. 516–23.

4. INCREMENTAL MAGNETIC PROPERTIES OF SILICON-IRON ALLOYS WITH PARTICULAR REFERENCE TO THE DESIGN OF AIR-GAPPED SMOOTHING CHOKES, R. O. Carter, D. L. Richards. *Proceedings*, Institution of Electrical Engineers, London, England, vol. 97, pt. II, Apr. 1950, pp. 199–214.

5. OPTIMUM AIR GAP FOR VARIOUS MAGNETIC MATERIALS IN CORES OF COILS SUBJECT TO SUPERPOSED DIRECT CURRENT, V. E. Legg. *AIEE Transactions*, vol. 64, Oct. 1945, pp. 709–12.

6. FINAL REPORT, RESEARCH AND DEVELOPMENT ON DESIGN METHOD FOR REACTORS, J. T. Ludwig. University of Minnesota, Minneapolis, Minn., contract no. DA36-039-SC-42573, Sept. 15, 1954, pp. 27–141.

7. A NEW METHOD FOR INDUCTOR DESIGN, J. T. Ludwig. *Proceedings*, Electronic Components Conference, Los Angeles, Calif., 1955, pp. 18–21.

8. TRANSFORMER LAMINATIONS, EM 3. Allegheny Ludlum Steel Corporation, Pittsburgh, Pa., fifth edition, 1952.

An Approach to Dynamic Optimizing Control of the Continuous Process

J. F. SANDELIEN
NONMEMBER AIEE

Synopsis: Optimum performance of the multivariable continuous process under transient conditions is considered. A generalized performance criterion is formulated in terms of the cost of operation. Assuming that the essential process behavior is known and that the major disturbance variables may be measured, the necessary conditions for minimum cost of operation are derived by applying calculus of variations in conjunction with the Lagrange multiplers technique. In certain typical cases of process behavior and cost criteria, the resulting necessary conditions reduce to a form which will yield an analytical solution. In particular, by certain approximations of the system behavior in the vicinity of an operating point, the necessary condition for optimum performance is expressed in terms of an inhomogenous Wiener-Hopf integral equation of the second kind and the solution is obtained by spectral factorization. In general, the conditions to be satisfied for minimum cost of operation appear as a set of nonlinear integral equations. A method for handling the general case is suggested.

THE CONTINUOUS process is defined as a system in which there is a continuous conversion of raw material into product. In the steady state, the process variables are constant with respect to time; they are functions, however, of the spatial co-ordinates of the processing plant. In the nonsteady or transient state, the continuous process is characterized by changes in the level of operation caused by disturbances such as variations in quality of raw materials, changes in demand, product specifications, environment, etc. Maintaining peak performance during transient as well as steady-state conditions is a problem of considerable importance and complexity, depending on the interrelationships between the process variables.

Fig. 1 represents a generalized configuration of a continuous process in which the variables are classified as:

1. Manipulated variables $\overline{m}_i$ which are subjected to direct control and provide the means for guiding the process along the path of optimum performance.
2. Disturbance variables $\overline{\vartheta}_j$ which are independent and include load- and set-point variables.
3. Output variables $\overline{q}_k$ which represent product quality, yield, etc.
4. State variables $\overline{y}_l$ representing the intermediate variables which serve to establish the relationships between the input and output variables.

To obtain peak performance, the optimum combination of variables must be determined according to some criterion

---

Paper 60-247, recommended by the AIEE Computing Devices Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., January 31–February 5, 1960. Manuscript submitted November 4, 1959; made available for printing January 20, 1960.

J. F. SANDELIEN, formerly with the Case Institute of Technology, Cleveland, Ohio, is now with the IBM Research Center, Yorktown Heights, N. Y.

This paper represents a thesis investigation conducted at Case Institute of Technology in partial fulfillment of the requirements for an M.S. degree in Instrumentation Engineering. Dr. I. Lefkowitz was responsible for guidance and stimulation in the work, which was carried out in conjunction with a sponsored research project in computer and optimizing control. The author wishes to acknowledge the present support of this area of research at Case Institute of Technology by the General Electric Company, the International Business Machines Corporation, Republic Steel Corporation, Thompson-Ramo-Wooldridge Inc., and the Westinghouse Electric Corporation.

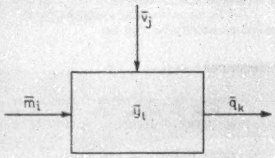


Fig. 1. Generalized configuration of a continuous process

$i=1,2,\ldots,n$
$j=1,2,\ldots,p$
$k=1,2,\ldots,r$
$l=1,2,\ldots,s$

of performance. A useful index or criterion is provided by the cost of production, which depends on such items as raw materials, energy, labor, by-product recovery, off-standard product, etc. It is assumed that the cost rate C' can be expressed as a function of the process variables:

$$C' = f_1(\bar{m}_i, \bar{q}_k, \bar{y}_s) \qquad (1)$$

where $i=1, 2, \ldots, n$; $k=1, 2, \ldots, r$; and $l=1, 2, \ldots, s$.

The techniques presently employed for the purpose of achieving minimum cost of production are listed and described briefly as follows:

1. *Model Methods.* Assuming that the steady-state relationships between the variables are known, the outputs $\bar{q}_k$ may be expressed in terms of the manipulated variables $\bar{m}_i$ and the disturbances $\bar{v}_j$. The necessary conditions for the steady-state optimum combination of values of the manipulated variables are obtained by differentiating equation 1 with respect to each of the n manipulated variables and equating the resulting partial derivatives to zero:

$$\frac{\partial C'}{\bar{m}_i} = 0, \quad i=1, 2, \ldots, n \qquad (2)$$

Implicit relationships may be handled by the introduction of Lagrangian multipliers. The necessary steps are often carried out as a continuous computer operation.

2. *Nonmodel Methods.* This technique is based in principle on a trial-and-error procedure which may be carried out with little or no knowledge about the relationships between the variables. First, the manipulated variables are altered according to a certain scheme. The resulting change in the performance is then evaluated relative to these alterations in order to determine the direction of further changes in the manipulated variables towards an optimum combination. This sequence repeats itself and may proceed continuously or in discrete steps.

These techniques, as presently applied, ignore process dynamics. The result is that operation during transient conditions cannot be optimum. In cases where disturbances occur at frequencies significant compared with the process dynamics, changes in the manipulated variables generated by an optimizing scheme that does not consider the dynamics may prolong a nonoptimum transient from one disturbance to the next. This means that the optimum steady state is never achieved, and the problem of optimizing during transient conditions becomes all-important. The optimum control of the continuous process under these dynamic conditions is the problem considered in this paper.

A mathematical formulation of the problem is provided by making the following stipulation:

1. It is assumed that through analysis, experimental observation, or self-checking, the steady-state relationships and dynamic functions for the system are known. (Self-checking is referred to as a method for determining system behavior from past and current information supplied by the process variables.[1])

2. It is assumed that the process is displaced from the steady state by action of one or more of the disturbance variables. The disturbances are assumed to be finite so that the process can be controlled to a new steady-state configuration corresponding to an optimum change in cost rate, designated C_s.

3. The initial steady-state values of the variables are used to provide a reference level for defining the variables during the transient period. Where the deviations of the variables from their respective reference values remain small during the transition period, certain approximations may be applied to the system equations.

The average cost of operation over an indefinite period of time is

$$\bar{c} = \lim_{T\to\infty} \frac{1}{T}\int_0^T f_1(\bar{m}_i, \bar{q}_k, \bar{y}_l)dt \qquad (3)$$

where $f_1(\bar{m}_i, \bar{q}_k, \bar{y}_l)$ is the cost rate given by equation 1. Now, defining the variables in terms of their steady-state reference values gives

$$\bar{m}_i = M_i + m_i$$
$$\bar{q}_k = Q_k + q_k$$
$$\bar{y}_l = Y_l + y_l$$
$$\bar{c} = C + c \qquad (4)$$

The bar refers to the total value of the variables, upper-case letters to the initial steady-state condition, and lower-case letters designate changes from the initial state.

Substituting into equation 3, the following equation is obtained:

$$\bar{c} = C + \lim_{T\to\infty} \frac{1}{T}\int_0^T f(m_i, q_k, y_l)dt \qquad (5)$$

where $f(m_i, q_k, y_l)$ is a function of the deviations in the variables.

The problem now reduces to finding the functions $m_i(t)$ that minimize the integral of equation 5. The optimizing conditions do not change if a constant

$$C_s = \lim_{T\to\infty} \frac{1}{T}\int_0^T C_s dt$$

is subtracted from the integral of equation 5. The new integrand, $f(m_i, q_k, y_l) - C_s$, converges to zero as T goes to infinity as a result of assumption 2, above. Thus the cost is minimized by minimizing the integral:

$$I = \int_0^\infty [f(m_i, q_k, y_l) - C_s]dt \qquad (6)$$

subject to the constraints imposed by the system equations. This formulation, together with the above stipulations, constitute the mathematical statement of the problem. (Equation 6 is a more convenient form than the last term of equation 5 from the standpoint of later manipulations.)

Related Problems

Calculus of variations provides a useful tool in the solution of engineering problems dealing with maximizing or minimizing quantities defined by a definite integral. Several applications are prominent in the area of automatic control referred to as "control design with specified criteria." A summary of these applications is presented here because of their resemblance to the present application problem.

Control design with specified criteria is discussed in references 2 and 3, where the objective is to guide a system from one essential operating level to another in an optimum fashion determined by some performance criterion. The constraints are expressed in terms of differential equations. The necessary conditions for optimum performance appear likewise as differential equations which are subjected to the required boundary conditions for solution. A stability consideration is necessary, once the system reaches the new operating level, in order to avoid its running away from the desired operating point.

This technique is successfully extended to the area of process control in reference 4 where the optimum control of a chemical batch process is considered. The product composition of the batch is guided along an optimum path from initial to final values by repetitive computer control action which is based on the solution of a set of differential equations satisfying given boundary conditions. Stability of the system, once the end point has been reached, is obviously of no consequence.

The analytic design of compensating elements for servomechanisms is treated extensively in reference 5. The performance index appears either as the integral square error, or the mean square error, depending on the input signals considered. The necessary condition for optimum performance, obtained by application of the Wiener theory of optimum filters, is expressed in the form of an integral equation. The present problem in its elementary configuration is analogous to developments in reference 5. The derivation of the necessary conditions on the basis of integral equations is a result of the problem formulation. The boundary conditions are an integral part of the problem and no separate considerations with regard to stability are required outside the criteria for optimum performance.

Derivation of the Necessary Conditions for Optimum Performance Using Generalized System Configuration

The relationships between the variables in the system, Fig. 1, are in general expressed by a set of consistent and independent nonlinear differential equations. It is reasonable to assume, however, that the steady-state relationships are of primary importance in the optimum control of the continuous process, and that a linear approximation of the system dynamics within a specified region about the operating point will cause negligible deviations from the theoretical optimum.

Although the minimizing conditions may be obtained with nonlinear dynamics included in the system equations, the form given by relations 8 in the following is considered adequate in most cases.

First, a generalized notation is adopted for the variables, q, y, and m:

$$\begin{array}{ll} q_k = w_k & k = 1, 2, \ldots, r \\ y_l = w_{r+l} & l = 1, 2, \ldots, s \\ m_i = w_{r+s+i} & i = 1, 2, \ldots, n \end{array} \tag{7}$$

Then the process behavior is expressed by the following set of equations:

$$\sum_{\alpha=0}^{N} a_\alpha \frac{d^\alpha w_u}{dt^\alpha} = H_u(w_z, v_j) \tag{8}$$

where

w_u is defined by relation 7: $u = 1, 2, \ldots, r, r+1, \ldots, r+s$
a_α = constants: $\alpha = 1, 2, \ldots, N$
t = time variable
$H_u(w_z, v_j)$ = general nonlinear functions of the arguments w_z and v_j, which in turn are functions of time.

$u = 1, 2, \ldots, r, r+1, \ldots, r+s$
$z = 1, 2, \ldots, r, r+1, \ldots, r+s, r+s+1, \ldots, r+s+n$
$j = 1, 2, \ldots, p$

Definition of the variables by equation 4 permits the transformation of equation 8 into an integral equation as follows: With Laplace transforms

$$\mathfrak{L}(w_u) = \frac{\mathfrak{L}[H_u(t)]}{\sum_{\alpha=0}^{N} a_\alpha s^\alpha} \tag{9}$$

where

$\mathfrak{L}$ = Laplace transform operator
s = Laplace transform variable

By the convolution theorem

$$w_u(t) - \int_{-\infty}^{\infty} H_u(t_1) h_u(t-t_1) dt_1 = 0 \tag{10}$$

where

$$h_u(t) = \mathfrak{L}^{-1} \frac{1}{\sum_{\alpha=0}^{N} a_\alpha s^\alpha}$$

is a known weighting function.

The integral of equation 6 is now minimized, subject to the constraint equations 10, by application of the method of Lagrange multipliers and the calculus of variations. The procedure is similar to that given in reference 6.

Equations 10 are multiplied by the arbitrary functions $\mu_u(t)$ and added to the integrand of equation 6:

$$I = \int_{-\infty}^{\infty} \left(f_s + \sum_{u=1}^{r+s} \mu_u \phi_u \right) dt \tag{11}$$

where

$f_s \equiv f(m_i, q_k, y_l) - C_s u(t)$; $u(t)$ = unit step at $t=0$
$\phi_u \equiv$ left-hand side of equation 10

Since the integrand of equation 11 is zero for $t<0$ (a result of definition 4), the lower limit of integration is extended to $-\infty$ for later convenience.

The following expressions are substituted into equation 11:

$$W_z(t) = w_z(t) + \epsilon \bar{w}_z(t); \quad z = 1, 2, \ldots, r+s+n \tag{12}$$

where $w_z(t)$ are the assumed minimizing functions; ϵ is an arbitrary parameter; and $\bar{w}_z(t)$ are some arbitrary realizable functions consistent with equation 9, and are such that $w_z(t) = 0$, for $t \leq 0$.

The resulting integral becomes a function of the parameter ϵ:

$$I(\epsilon) = \int dt \left[f(W_z) - u(t) C_s + \sum_{u=1}^{r+s} \mu_u W_u - \sum_{u=1}^{r+s} \mu_u \int dt_1 H_u(t_1) h_u(t-t_1) \right] \tag{13}$$

(The limits of integration, where not otherwise designated, are from minus to plus infinity.) Since $W_z = w_z$ for $\epsilon = 0$, a necessary condition for the minimizing functions is that the derivative of the integral with respect to ϵ vanish for $\epsilon = 0$:

$$\frac{dI(\epsilon)}{d\epsilon} = \int dt \left[\sum_{z=1}^{r+s+n} \frac{\partial f}{\partial W_z} \frac{\partial W_z}{\partial \epsilon} + \sum_{u=1}^{r+s} \mu_u \frac{\partial W_u}{\partial \epsilon} - \sum_{u=1}^{r+s} \mu_u \int dt_1 \times \sum_{z=1}^{r+s+n} \frac{\partial H_u(t_1)}{\partial W_z} \frac{\partial W_z}{\partial \epsilon} h_u(t-t_1) \right] \tag{14}$$

where

$$\frac{\partial W_z}{\partial \epsilon} = \bar{w}_z \tag{15}$$

For

$$\epsilon = 0: \quad \frac{\partial f}{\partial W_z} = \frac{\partial f}{\partial w_z}$$

$$\text{and } \frac{\partial H_u(t_1)}{\partial W_z} = \frac{\partial H_u(t_1)}{\partial w_z}$$

$$\text{and } \frac{dI(\epsilon)}{d\epsilon} = 0$$

Integrating term by term in the second integral, interchanging the order of integration and applying relationships 15:

$$\int dt_1 \left[\sum_{z=1}^{r+s+n} \frac{\partial f}{\partial w_z} \bar{w}_z(t_1) + \sum_{u=1}^{r+s} \mu_u(t_1) \bar{w}_u(t_1) - \sum_{u=1}^{r+s} \sum_{z=1}^{r+s+n} \bar{w}_z(t_1) \int dt\, \mu_u(t) \times \frac{\partial H_u(t_1)}{\partial w_z} h_u(t-t_1) \right] = 0 \tag{16}$$

Collecting terms with the same limits of summation and factoring out the variation function $\bar{w}_z(t)$:

$$\int dt_1\left\{\sum_{u=1}^{r+s}\bar{w}_u(t_1)\left[\frac{\partial f}{\partial w_u}+\mu_u(t_1)-\right.\right.$$

$$\left.\sum_{x=1}^{r+s}\int dt\,\mu_x(t)\,\frac{\partial H_x(t_1)}{\partial w_u}\,h_x(t-t_1)\right]+$$

$$\sum_{y=r+s+1}^{r+s+n}\bar{w}_y(t_1)\left[\frac{\partial f}{\partial w_y}-\sum_{x=1}^{r+s}\times\right.$$

$$\left.\left.\int dt\mu_x(t)\,\frac{\partial H_x(t_1)}{\partial w_y}\,h_x(t-t_1)\right]\right\}=0 \quad (17)$$

Defining

$$F(t_1)=f+\sum_{u=1}^{r+s}\mu_u(t_1)w_u-$$

$$\sum_{x=1}^{r+s}\int dt\,\mu_x(t)H_x(t_1)h_x(t-t_1) \quad (18)$$

will reduce equation 17 to the following form:

$$\int dt_1\left[\sum_{u=1}^{r+s}\bar{w}_u(t_1)\,\frac{\partial F}{\partial w_u}+\right.$$

$$\left.\sum_{y=r+s+1}^{r+s+n}\bar{w}_y(t_1)\frac{\partial F}{\partial w_y}\right]=0 \quad (19)$$

Because of the constraints specified by equation 10, only n of the variation functions are independent. However, the r+s arbitrary functions $\mu_1(t_1)$, $\mu_2(t_1)$ $\ldots$, $\mu_{r+s}(t_1)$ may be so chosen as to make r+s of the partial derivatives of equation 19 vanish. If, as an example, the vanishing terms are $\partial F/\partial w_u$, $u=1, 2, \ldots, r+s$; the remaining variation functions $\bar{w}_z$ may be independent and arbitrary within the defining equation 12. In particular, for $t_1>0$, these functions may be nonzero. Consequently, that equation 19 may be satisfied, the remaining partial derivatives must also vanish. Thus, by using relationships 7, the necessary conditions for the minimizing function may be expressed as

$$\frac{\partial F}{\partial w_z}=0,\ z=1, 2, \ldots, r+s+n;\ \text{for } t_1\geq 0 \quad (20)$$

These equations are solved in conjunction with the constraint equations 10 to determine the manipulated variables for optimizing performance.

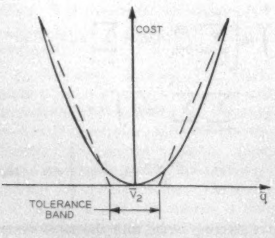


Fig. 2. Production cost versus deviation $\bar{q}$ in product quality

Optimum Performance Applied to Typical System Configurations

PERFORMANCE CRITERION

The general criterion for optimum performance, expressed in terms of the cost of production, must be evaluated by management in each individual case to establish the functional relationships of equation 1. The purpose here is not to discuss the different aspects of such evaluation, but to select a typical form of the cost function which is likely to satisfy a variety of process operations, and which will keep the complexity of the problem and its solution within the scope of practical utility.

The majority of operations demand a product quality within certain tolerances. Deviation from the given specifications may result in loss of product or a product of inferior quality which requires recycling, blending operations, etc.

The resulting increase in the cost of production may take the form shown graphically by the broken lines in Fig. 2, where the quality $\bar{q}$ is measured from a desired reference or set point $\bar{v}_2$. For mathematical convenience, this cost-quality relationship is approximated by the solid-line parabola in the same figure, and the cost of production resulting from a deviation in product quality is given by $C_2\bar{q}^2$, where C_2 is a constant.

The cost of raw material is considered proportional to the rate of consumption by a factor C_1. Assuming that the product flow rate is fixed, the cost equation for the system shown in Fig. 3 is

$$f_1(\overline{m}, \bar{q})=C_1\overline{m}+C_2\bar{q}^2 \quad (21)$$

This is the basic cost relationship used in the following system configurations.

SYSTEM WITH ONE SECOND-ORDER NONLINEARITY

Configuration

A system performing according to cost equation 21, and otherwise exhibiting linear characteristics, is referred to as a system with one second-order nonlinearity. A typical block diagram is shown in Fig. 4, where the product quality q is a function of a load variable v_1, a set point variable v_2, and a manipulated variable m. (Set point is referred to here as the desired value of the product quality.) The variables are defined according to equation 4. The system transfer functions are given by $K_1G(s)$ and $K_2N(s)$ where K_1 and K_2 are steady-state gain constants.

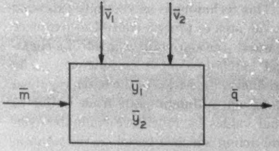


Fig. 3. Typical process configuration

The cost function, equation 21, is expressed in terms of an initial steady-state cost and a deviation, by application of relations 4:

$$f_1(\overline{m}, \bar{q})=C+f(m, q) \quad (22)$$

where $C=C_1M+C_2Q^2$, and $f(m, q)=C_1m+2C_2Qq+C_2q^2$.

Considering the case of fixed set point, the constraint equations expressed in the frequency domain are

$$q(s)=-y_1(s)-y_2(s)$$
$$y_1(s)=K_1G(s)m(s)$$
$$y_2(s)=K_2N(s)v_1(s) \quad (23)$$

In the time domain:

$$q(t)+\int[y_1(t_1)+y_2(t_1)]\delta(t-t_1)dt_1=0$$
$$y_1(t)-K_1\int m(t_1)g(t-t_1)dt_1=0$$
$$y_2(t)-K_2\int v_1(t_1)n(t-t_1)dt_1=0 \quad (24)$$

where

$\delta(t)$ = Dirac delta function defined, as

$$\delta(t)=0,\ \text{for } t\neq 0$$

$$\int_{-\infty}^{\infty}\delta(t)dt=1$$

Equations 22 and 24 completely define the over-all system behavior from the standpoint of cost of operation.

Solution

The necessary conditions for determination of the manipulated variable m

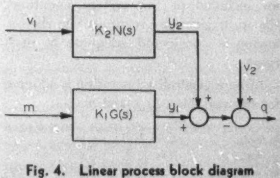


Fig. 4. Linear process block diagram

for optimum performance are obtained by application of equations 18 and 20. The relationships are

$$\begin{aligned} &w_1=q; && H_1=-(y_1+y_2); && h_1=\delta \\ &w_2=y_1; && H_2=K_1m; && h_2=g \\ &w_3=y_2; && H_3=K_2v_1; && h_3=n \\ &w_4=m && && \end{aligned} \tag{25}$$

Equation 18 yields

$$\begin{aligned} F(t_1)=&\,C_1m(t_1)+2C_2Qq(t_1)+C_2q^2(t_1) \\ &+\mu_1(t_1)q(t_1)+\int dt\ \mu_1(t)[y_1(t_1)+y_2(t_1)]\delta(t-t_1) \\ &+\mu_2(t_1)y_1(t_1)-K_1\int dt\ \mu_2(t)m(t_1)g(t-t_1) \\ &+\mu_3(t_1)y_2(t_1)-K_2\int dt\ \mu_3(t)v_1(t_1)n(t-t_1) \end{aligned} \tag{26}$$

for $t_1\geq 0$, with the use of relations 20:

$$\frac{\partial F}{\partial q}=2C_2Q+2C_2q(t_1)+\mu_1(t_1)=0$$

$$\frac{\partial F}{\partial y_1}=\mu_2(t_1)+\mu_1(t_1)=0$$

$$\frac{\partial F}{\partial y_2}=\mu_3(t_1)+\mu_1(t_1)=0$$

$$\frac{\partial F}{\partial m}=C_1-K_1\int dt\ \mu_2(t)g(t-t_1)=0 \tag{27}$$

which, combined with the constraint equations 24, reduces to

$$\begin{aligned} &K_1\int dt\ g(t-t_1)\int dt_2m(t_2)g(t-t_2)+ \\ &\qquad K_2\int dt\ g(t-t_1)\int dt_2v_1(t_2)n(t-t_2)- \\ &\qquad\quad Q\int dt\ g(t-t_1)+\frac{C_1}{2C_2K_1}=0 \text{ for } t_1\geq 0 \end{aligned} \tag{28}$$

Finally, interchanging the order of integration and introducing translation function

$$\begin{aligned} I_{gg}(w)&=\int dt\ g(t)g(t+w) \\ I_{gn}(w)&=\int dt\ g(t)n(t+w) \end{aligned} \tag{29}$$

gives the following equation:

$$\begin{aligned} &K_1\int dt_2m(t_2)I_{gg}(t_1-t_2)+ \\ &\qquad K_2\int dt_2v_1(t_2)I_{gn}(t_1-t_2)- \\ &\qquad\quad Q\int dt\ g(t-t_1)+\frac{C_1}{2C_2K_1}=0 \text{ for } t_1\geq 0 \end{aligned} \tag{30}$$

This is a nonhomogenous Wiener-Hopf integral equation of the first kind which is solved for $m(t)$ by the method of spectral factorization outlined in reference 7. The solution is obtained in the frequency domain by application of the bilateral Laplace transform to equation 30:

$$\begin{aligned} &\left[\frac{C_1}{2C_2K_1}-QG(-s)\right]2\pi\delta(js)+ \\ &\qquad K_1m(s)I_{gg}(s)+K_2v_1(s)I_{gn}(s)=A(s) \end{aligned} \tag{31}$$

where $A(s)$ is analytic in the entire LHP (left-half plane) since the inverse transform of $A(s)$ is zero for $t_1\geq 0$.

$$\bar{\mathfrak{L}}\int dt\ g(t-t_1)=2\pi\delta(js)G(-s) \tag{32A}$$

(See Appendix I.)

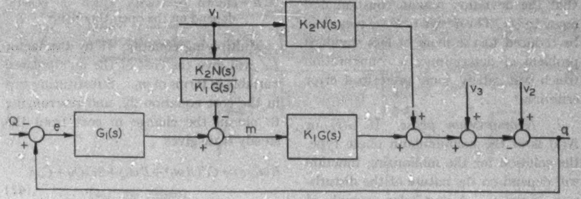


Fig. 5. **Optimum control of system with one second-order nonlinearity (minimum phase elements)**

$\bar{\mathfrak{L}}$ is the bilateral Laplace transform operator.

$\bar{\mathfrak{L}}(1)=2\pi\delta(js)$ is a valid transform for $Re(s)=0$ (reference 8), and

$$\begin{cases} \delta(js)=0 \text{ for } js\neq 0 \\ \displaystyle\int_{-\infty}^{\infty}\delta(js)d(js)=1 \end{cases} \tag{32B}$$

Condition 32B implies that the first term of equation 31 vanishes for all values of $s\neq 0$. When $s=0$, $G(-s)=1$ for all stable systems, and considerations of steady-state optimum performance shows that:

$$Q=\frac{C_1}{2C_2K_1} \tag{33}$$

(see Appendix II). Thus the first term of equation 31 vanishes identically for all values of s.

The solution of equation 31 is initiated by factorizing:

$$I_{gg}(s)=\gamma^+(s)\gamma^-(s) \tag{34}$$

where $\gamma^+(s)$ includes all factors with poles and zeros in LHP, and $\gamma^-(s)$ includes all factors with poles and zeros in RHP (right-half plane).

Then

$$K_1m(s)\gamma^+(s)+\frac{K_2v_1(s)I_{gn}(s)}{\gamma^-(s)}=\frac{A(s)}{\gamma^-(s)} \tag{35}$$

where $A(s)/\gamma^-(s)$ contains poles in the RHP only, since $A(s)$ is analytic in the entire LHP; and $K_1m(s)\gamma^+(s)$ contains poles in the LHP only.

Consequently, the nonsteady-state solution is

$$K_1m(s)\gamma^+(s)=-\left[\frac{K_2v_1(s)I_{gn}(s)}{\gamma^+(s)}\right]_+ \tag{36}$$

where []+ denotes partial fraction expansion terms for LHP poles of the bracketed expression.

If $[K_2v_1(s)I_{gn}(s)]/[\gamma^-(s)]$ is not rational, the left side of equation 36 is obtained by taking the inverse bilateral transform by complex inversion of the bracketed term and then performing a unilateral transformation of the resulting time function.

Implementation

The nature of the solution depends on whether the transfer functions $G(s)$ and $N(s)$ are minimum or nonminimum phase:

1. *Minimum phase.* From Appendix I:

$$I_{gg}(s)=G(-s)G(s) \tag{37A}$$

$$I_{gn}(s)=G(-s)N(s) \tag{37B}$$

Further:

$$\gamma^+(s)=G(s) \text{ and } \gamma^-(s)=G(-s) \tag{38}$$

and, since

$K_2N(s)v_1(s)$

has poles in the LHP only:

$$m(s)=-\frac{K_2}{K_1}\frac{N(s)}{G(s)}v_1(s) \tag{39}$$

A control system satisfying this relationship would be of the form shown in Fig. 5. Considering minimum phase elements, it is theoretically possible to achieve the ideal optimum by exact compensation in the variable m for a change in the variable v_1. This will maintain the deviation constant at the steady-state optimum value given by equation 33. In practice, however, exact compensation is not possible, due to saturation, inaccurate matching of parameters, etc. In addition, load variables, such as v_3 in Fig. 5 may exist which cannot be adequately compensated because of difficulties in measurement, etc. To take care of these imperfections and to provide the necessary bias for optimum steady-state operation, negative feedback is provided. In this particular example, the optimum deviation is a constant for the steady-state as well as the transient period, a result which perhaps could have been anticipated on the basis of steady-state relationships.

The purpose of the $G_1(s)$ function in Fig. 5 is to reduce the actuating signal $e(t)$ to zero, where

$$-e(t)=Q-q \tag{40}$$

This satisfies the optimizing condition

that the deviation remain constant and equal to Q. The design of $G_1(s)$ may then be reduced to the more or less standard problem of determining a configuration which will satisfy some prescribed error criterion.[5]

2. *Nonminimum phase.* If $G(s)$ or $N(s)$ is of the nonminimum phase type, the solution for the minimizing function will depend on the nature of the disturbance signal v_2. A particular example of interest is the presence of dead time in the control loop. Here, exact compensation is physically unrealizable; however, equation 36 may be solved for the best realizable compensation for a given disturbance function.

SYSTEM WITH TWO SECOND-ORDER NONLINEARITIES

Configuration

The development in the preceding section has severe limitations with regard to application to optimizing continuous processes because the configuration implies that the system is completely linear. The analysis is based on deviations from a steady-state operating point, which in conventional control problems permits the usual linearizing technique. This procedure of eliminating troublesome nonlinearities, however, defeats the purpose of optimizing control, and the case where nonlinearities appear in the process equations must be considered.

The example of the preceding section does provide some useful information with regard to the mathematical technique required, and the type of nonlinearities in the system equations which will permit an analytical solution by established mathematical procedures. The following approximations of the general system behavior is based on these observations.

The nonlinear process equations are first linearized in a conventional manner. Realizing the necessity for including the nonlinear gain characteristics, these are established and approximated by the first- and second-order terms in a power series expansion about the operating point. A mathematical model of the actual system is then constructed, as shown in Fig. 6. The model consists of an element containing the linearized dynamics with unity gain at zero frequency and a second element representing the approximated nonlinear gain in terms of deviations in the variables from a steady state:

$$m=Am_1^2+Bm_1 \tag{41}$$

where

m = manipulated variable
m_1 = a fictitious variable
A, B = known constants which in general depend on the operating point

Multiplying equation 41 by the factor C_1' will give the cost of the manipulated variable in terms of m_1. Substituting this in the cost equation 22, and rearranging to obtain the change in cost from the steady state gives

$$f(m_1, q)=C_1'(Am_1^2+Bm_1)+2C_2Qq+C_2q^2 \tag{42}$$

The constraint equations have been linearized at the expense of adding a second-order term to the cost equation, thus retaining the essential nonlinear gain characteristics. This is referred to here as a system with two second-order nonlinearities. The problem can be solved analytically for $m_1(t)$, which is substituted in equation 41 to obtain the manipulated variable $m(t)$.

To simplify the mathematical manipulations and to ease the interpretation of the final solution, the system configuration is changed to correspond to the process represented by Fig. 4, and having the following cost function (refer to equation 21):

$$\begin{aligned} f_1(\overline{m}, \bar{q})&=C_1\overline{m}^2+C_2\bar{q}^2\\ &=C+f(m, q) \end{aligned} \tag{43}$$

where

$$C=C_1M^2+C_2Q^2$$

$$f(m, q)=2C_1Mm+C_1m^2+2C_2Qq+C_2q^2 \tag{44}$$

Equation 44 contains the same terms as equation 42 and the procedure for obtaining a solution is identical.

Solution

Equations 23, 24, and 25 remain valid, and the optimizing conditions are obtained by use of equations 18 and 20:

$$\begin{aligned} F(t_1)={}&2C_1Mm(t_1)+C_1m^2(t_1)+2C_2Qq(t_1)+\\ &C_2q^2(t_1)+\mu_1(t_1)q(t_1)+\int dt\,\mu_1(t)\times\\ &[y_1(t_1)+y_2(t_1)]\delta(t-t_1)+\mu_2(t_1)y_1(t_1)-\\ &K_1\int dt\,\mu_2(t)m(t_1)g(t-t_1)+\mu_3(t_1)y_2(t_1)-\\ &K_2\int dt\,\mu_3(t)v_1(t_1)n(t-t_1) \end{aligned} \tag{45}$$

and for $t_1 \geq 0$

$$\frac{\partial F}{\partial q}=2C_2Q+2C_2q(t_1)+\mu_1(t_1)=0$$

$$\frac{\partial F}{\partial y_1}=\mu_2(t_1)+\mu_1(t_1)=0$$

$$\frac{\partial F}{\partial y_2}=\mu_3(t_1)+\mu_1(t_1)=0$$

$$\frac{\partial F}{\partial m}=2C_1M+2C_1m(t_1)-K_1\int dt\,\mu_2(t)g(t-t_1)=0 \tag{46}$$

which, combined with the constraint equations 24, reduces to

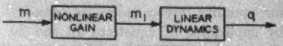


Fig. 6. Mathematical model of nonlinear process

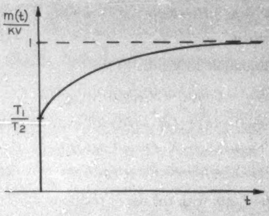


Fig. 7. Optimum response to step disturbance for system with two second-order nonlinearities

$$\begin{aligned} C_1M-C_2K_1Q\int dt\,g(t-t_1)+C_2K_1^2\int dt\times\\ g(t-t_1)\int dt_2m(t_2)g(t-t_2)+\\ C_2K_1K_2\int dt\,g(t-t_1)\int dt_2v_1(t_2)n(t-t_2)+\\ C_1m(t_1)=0 \text{ for } t_1\geq 0 \end{aligned} \tag{47}$$

Interchanging the order of integration and introducing translation function notation, obtains the following nonhomogenous Wiener-Hopf integral equation of the second kind:

$$\begin{aligned} C_1M-C_2K_1Q\int dt\,g(t-t_1)+C_2K_1^2\int dt_2m(t_2)\times\\ I_{gg}(t_1-t_2)+C_2K_1K_2\int dt_2v_1(t_2)I_{gn}(t_1-t_2)\\ +C_1m(t_1)=0 \text{ for } t_1\geq 0 \end{aligned} \tag{48}$$

The method of solution outlined in reference 7 is extended to include integral equations of the second kind as follows (a more extensive account is given in reference 9): the bilateral Laplace transform of equation 48 is

$$\begin{aligned} [C_1M-C_2K_1QG(-s)]2\pi\delta(js)+\\ [C_2K_1^2I_{gg}(s)+C_1]m(s)+\\ C_2K_1K_2v_1(s)I_{gn}(s)=A(s) \end{aligned} \tag{49}$$

Optimum steady-state relationships derived in Appendix II show that

$$C_1M=C_2K_1Q \tag{49A}$$

Then, in view of relation 32B, the first term of equation 49 vanishes identically. By factorization

$$[C_2K_1^2I_{gg}(s)+C_1]=\gamma^+(s)\gamma^-(s) \tag{50}$$

The solution is obtained as before:

$$m(s)\gamma^+(s)=-\left[\frac{C_2K_1K_2v_1(s)I_{gn}(s)}{\gamma^-(s)}\right]_+ \tag{51}$$

In this case no simplification of the solution formula for minimum phase elements is possible since $\gamma^-(s)\neq G(-s)$. In general, the solution depends on the nature of the disturbance (v_1).

One particular solution is derived in Appendix III for the simple case where

Vol.
Pt. 1
196

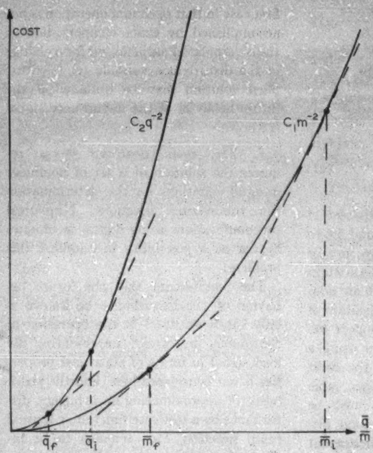


Fig. 8 (left). Cost relationship for system with two second-order nonlinearities

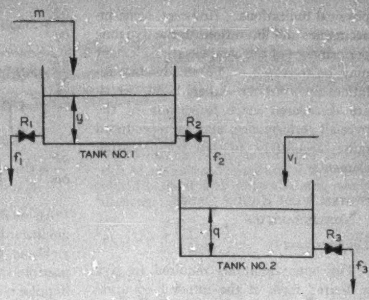


Fig. 10 (right). Hydraulic analog of system with higher-order nonlinearities

$$G(s)=N(s)=\frac{1}{1+Ts} \tag{52}$$

and

$$v_1(s)=V/s \tag{53}$$

Here T is the time constant of the first order elements and V is the magnitude of the step disturbance. Then

$$m(s)=-K\frac{1+T_1s}{1+T_2s}v_1(s) \tag{54}$$

where $T_2>T_1$ and K, T_1 and T_2 are known constants determined by the values of system gains (K_1 and K_2) and cost coefficients (C_1 and C_2).

The corresponding time function is

$$m(t)=-KV\left[\left(\frac{T_1}{T_2}-1\right)\epsilon^{-t/T_2}+1\right] \tag{55}$$

which is plotted in Fig. 7.

To evaluate this solution, the steady-state cost characteristics (see equation 43) are displayed in Fig. 8. In any one optimum steady state the corresponding values of the manipulated variable ($\bar{m}$) and the quality ($\bar{q}$) are located on the cost curves of Fig. 8 in such a manner that the incremental costs, identified by the slope of each curve, are the same. When the process is subjected to a step change in the disturbance variable (see equation 53) the initial and final optimum steady-state values for the variables ($\bar{m}$), and ($\bar{q}$) are given in Fig. 8 by ($\bar{m}_i$, $\bar{q}_i$) and ($\bar{m}_f$, $\bar{q}_f$) respectively. The optimum path between these steady states may be explained qualitatively as follows: Assume that $\bar{m}(t)$ is changed directly to the final value in the form of a step change. As a result, $\bar{q}(t)$ is reduced exponentially to the new steady state, and from Fig. 8 it is observed that the incremental cost of $\bar{q}(t)$ is higher than that of $\bar{m}(t)$ at all times during the transient. Thus, it is clear that an increased rate of reduction in $\bar{q}(t)$ at the expense of a slower reduction in $\bar{m}(t)$ will improve the transient performance from the standpoint of minimum cost. This agrees with the analytical solution for $\bar{m}(t)$ given by equation 55 and shown in Fig. 7.

Implementation

The analytical solution for the optimizing function using system configurations stipulated in this section is limited by the requirement that the disturbances be known. A control system for a process where the disturbance (v_1) enters the system in discrete steps, and otherwise performs according to equation 23, would be of the form shown in Fig. 9. Negative feedback is provided in order to maintain desired operating conditions in spite of extraneous load disturbances (such as v_2) which cannot be accounted for because of difficulties in measurement, etc. The desired value of the variable q is computed (see equation 76) in Appendix III on the basis of the essential disturbance (v_1). Any deviation (e) from this value during the course of operation is reduced to zero by the action of the controller element $G_1(s)$.

Since all relationships of Fig. 9 are linear, the principle of superposition is valid, and the solution as given by equation 54 can be applied to step disturbances taking place under nonsteady-state conditions. This may be extended to the system of Fig. 6 if the coefficients K, T_1, and T_2 in the solution formula 54 are adjusted to correspond to the level of operation at at which the disturbance takes place. The extent of the approximation involved depends largely on the procedure for adjusting the coefficients.

The matching of an actual compensating element to a given transfer function has

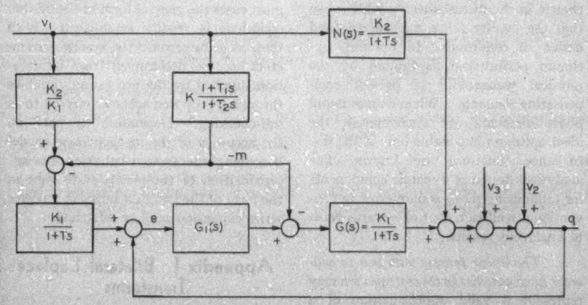


Fig. 9. Optimum control of system with two second-order nonlinearities subjected to step disturbances

physical limitations. However, some inaccuracies and limitations in the dynamic performance of the compensating element may be tolerated. The steady-state relationships on the other hand, which are considered more important in the over-all performance, may be reproduced quite accurately in the compensating elements.

SYSTEM WITH HIGHER-ORDER NONLINEARITIES

Configuration

The approximations required to give a desired form of the integral equation for analytical solution, as outlined in the preceding section, may not be feasible in certain cases.

In order to investigate the possibility of using the general nonlinear equations 8, a hypothetical process in the form of a hydraulic analog, shown in Fig. 10, is considered. The system is assumed to be completely linear except for a nonlinear resistance (R_2). The head-flow relationships are

$$f_1=y/R_1;\ f_2=\sqrt{y}/R_2;\ f_3=q/R_3 \qquad (56)$$

where f_1, f_2, f_3, y, and q are variables identified by Fig. 10, and R_1, R_2, and R_3 constants.

Then the constraint equations expressed in terms of the total value of the variables, are

$$E_1\frac{dy}{dt}+\frac{y}{R_1}=m-\frac{\sqrt{y}}{R_2}$$

$$E_2\frac{dq}{dt}+\frac{q}{R_3}=v_1+\frac{\sqrt{y}}{R_3} \qquad (57)$$

where E_1 and E_2 are the capacitances of tanks number 1 and 2 respectively.

In integral equation form, the constraints are (setting all constants equal to unity for convenience)

$$y-\int(m-\sqrt{y})g_1(t-t_1)dt_1=0$$
$$q-\int(v_1+\sqrt{y})g_2(t-t_1)dt_1=0 \qquad (58)$$

where $g_1(t)=g_2(t)=\mathfrak{L}^{-1}\,1/(s+1)$.

The cost function is the same as before:

$$f(m,\, q)=C_1m+C_2q^2 \qquad (59)$$

Possible Solution and Implementation

The necessary conditions for the optimizing function are obtained by equations 18 and 20:

$$\begin{aligned}F(t_1)=C_1m(t_1)+C_2q^2(t_1)+\mu_1(t_1)q(t_1)-\\ \int dt\ \mu_1(t)[v_1(t_1)+\sqrt{y(t_1)}]g_2(t-t_1)+\\ \mu_2(t_1)y(t_1)-\int dt\mu_2(t)[m(t_1)-\\ \sqrt{y(t_1)}]g_1(t-t_1)\end{aligned} \qquad (60)$$

and, for $t_1\geq 0$:

$$\frac{\partial F}{\partial q}=2C_2q(t_1)+\mu_1(t_1)=0$$

$$\frac{\partial F}{\partial y}=\mu_2(t_1)+\frac{1}{2}\int dt\,\frac{\mu_2(t)}{\sqrt{y(t_1)}}\,g_1(t-t_1)-\frac{1}{2}\int dt\,\frac{\mu_1(t)}{\sqrt{y(t_1)}}\,g_2(t-t_1)=0$$

$$\frac{\partial F}{\partial m}=C_1-\int dt\mu_2(t)g_1(t-t_1)=0 \qquad (61)$$

An analytical treatment of the set of nonlinear integral equations 61 and 58, to yield an explicit solution, may not be possible. Because the equations contain impulse response functions which are nonzero for negative time, the application of analog computer techniques appear remote, and a numerical solution using a digital computer seems to be the most promising possibility. The proper value of the manipulated variable must be available immediately upon a load disturbance, which requires a certain amount of precomputation of the response function for typical conditions. It appears that the feasibility of this procedure willl depend in particular on the type of disturbances encountered.

Conclusions

The problem of dynamic optimizing control of the continuous process is formulated and treated in considerable generality. The necessary conditions for optimum performance are derived and applied to system configurations of increasing degrees of complexity. The results may be summarized as follows:

1. *The linear process with one second-order nonlinearity in the cost equation*, and identified by minimum phase transfer functions, theoretically has no limitations in achieving the ideal optimum. The solution calls for a perfect compensation in the manipulated variable (m) for a change in the disturbance variable (v) so that the quality (q) is maintained fixed under all conditions. In practice, optimum performance is limited by the physical realizability of perfect compensating elements. When nonminimum phase elements are encountered, the ideal optimum is possible only if the disturbance functions are known. The analytical design of dynamic components for minimizing mean square error in conventional feedback control systems leads to analogous results.

2. *The linear process with two second-order nonlinearities in the cost equation* may represent a valid approximation of a general nonlinear configuration. The results are significantly different from the first case in that optimum operation is not accomplished by exact compensation in the manipulated variable (m) for a change in the disturbance variable (v). An explicit solution may be obtained if the future behavior of the disturbance signal is known.

3. *The general nonlinear process* requires the solution of a set of nonlinear integral equations for the determination the minimizing functions. Numerical precomputations using digital techniques appear as a possibility in handling this problem.

The requirement that the future behavior of the disturbance be known is only rarely satisfied in the operation of continuous processes. Representing the load signal in terms of statistical properties is not considered here, and the possibility of approximating an arbitrary disturbance by a staircase function, a limited ramp function, etc., remains to be investigated. In any case, the implication is that the ideal optimum cannot in general be achieved. However, proper design of compensating elements (or equivalent computer actions) based on typical load signals and operated on by the actual random disturbance to obtain the manipulated variable, will enable the system to seek an optimum steady state under any condition. Further, the transient performance is likely to be close to optimum for disturbances that do not deviate greatly from the design signal. Thus, continuous automatic adjustments toward an optimum combination of variables at any level of operation, and regardless of the magnitude of load fluctuations, appear possible. This means, of course, that conventional equipment for minimizing the effects of large load changes, such as storage tanks, surge vessels, etc., may be eliminated.

The technique developed here assumes that the system behavior is known. In most cases the general form of the system equations is readily established. With these as a framework, the system parameters may be determined from information supplied by the processing variables through a proposed scheme referred to as self-checking.[1] Depending in part on the accuracy of the mathematical model describing the system behavior, the approximation of the steady-state gain in the form of the two first terms in a power series expansion may be adequate.

Appendix I. Bilateral Laplace Transforms

The following particular transforms are used:

(a)

$$\overline{\mathfrak{L}}\int_{-\infty}^{\infty} dt\, g(t-t_1)=I_{g1}(s)=2\pi\delta(js)G(-s) \qquad (32\text{A})$$

which is obtained as follows, starting with the definition

$$I_{g1}(s)=\int_{-\infty}^{\infty} dt_1\, \epsilon^{-st_1}\int_{-\infty}^{\infty} dt\, g(t-t_1) \qquad (62)$$

Interchanging the order of integration and introducing the variable gives

$$t_2=t-t_1 \qquad (63)$$

Then,

$$I_{g1}(s)=\int dt\, \epsilon^{-st}\int dt_2\, \epsilon^{st_2}g(t_2) \qquad (64)$$

(Where the limits of integration, if not otherwise designated, are from minus to plus infinity) and equation 32(A) follows, since

$$\int dt\epsilon^{-st}=\overline{\mathfrak{L}}(1)=2\pi\delta(js) \qquad (65)$$

by reference 8.

(b)

$$\overline{\mathfrak{L}}[I_{gg}(w)]=I_{gg}(s)=G(-s)G(s) \qquad (37\text{A})$$

which is derived as follows:

$$I_{gg}(s)=\int dw\, \epsilon^{-sw}\int dt\, g(t)g(t+w) \qquad (66)$$

Interchanging the order of integration and introducing:

$$w_1=t+w \qquad (67)$$

Then,

$$I_{gg}(s)=\int dt\, g(t)\epsilon^{st}\int dw_1 g(w_1)\epsilon^{-sw_1} \qquad (68)$$

and equation 37(A) follows immediately

(c)

$$\overline{\mathfrak{L}}[I_{gn}(w)]=I_{gn}(s)=G(-s)N(s) \qquad (37\text{B})$$

This is obtained by the same procedure used for transform equation 37(A).

Appendix II. Derivation of Steady-State Optimum

System With One Second-Order Nonlinearity

From relationships 22, the steady-state cost rate is

$$C=C_1M+C_2Q^2 \qquad (69)$$

where the variables are defined according to equation 4.

The steady-state relationship (see Fig. 4)

$$Q=V_2-K_1M-K_2V_1 \qquad (70)$$

is substituted in equation 69, and the condition for minimum steady-state cost is obtained by differentiating the resulting equation with respect to the manipulated variable M. This yields:

$$M=\frac{V_2-K_2V_1}{K_1}-\frac{C_1}{2C_2K_1^2} \qquad (71)$$

and

$$Q=\frac{C_1}{2C_2K_1} \qquad (33)$$

System With Two Second-Order Nonlinearities

The steady-state cost rate is

$$C=C_1M^2+C_2Q^2 \qquad (44)$$

Substituting for Q from equation 70 and differentiating the cost rate equation 44 with respect to M, yield the following conditions:

$$M=\frac{C_2K_1}{C_1+C_2K_1^2}\,(V_2-K_2V_1) \qquad (72)$$

and

$$C_1M=C_2K_1Q \qquad (49\text{A})$$

Appendix III. Derivation of the Explicit Solution for a Particular System Configuration

The dynamics of a system with two second-order nonlinearities are given by the transfer functions (see Fig. 4)

$$G(s)=N(s)=\frac{1}{1+Ts} \qquad (52)$$

Then, from Appendix I

$$I_{gg}(s)=I_{gn}(s)=G(-s)G(s)=\frac{1}{(1+Ts)(1-Ts)} \qquad (73)$$

and equation 50 becomes

$$\frac{C_2K_1^2}{(1+Ts)(1-Ts)}+C_1=\gamma^+(s)\gamma^-(s) \qquad (74)$$

where

$$\gamma^+(s)=\sqrt{C_1}\,\frac{b+s}{a+s}$$

$$\gamma^-(s)=\sqrt{C_1}\,\frac{b-s}{a-s}$$

$$a=\frac{1}{T}$$

$$b=\frac{1}{T}\sqrt{\frac{C_2K_1^2+C_1}{C_1}}$$

For a step change in the disturbance variable

$$v_1(s)=V/s \qquad (53)$$

and the bracketed term of equation 51 is expanded in partial fractions as follows:

$$\frac{I_{gn}(s)v_1(s)}{\gamma^-(s)}=\frac{V}{T^2\sqrt{C_1}}\,\frac{1}{s(a+s)(b-s)}=\frac{V}{T^2\sqrt{C_1}}\left[\frac{1/ab}{s}-\frac{1/a(a+b)}{a+s}+\frac{1/b(a+b)}{b-s}\right] \qquad (75)$$

Eliminating the term with the RHP pole from equation 75 and substituting in equation 51 obtains the transform of the manipulated variable with some algebraic manipulations:

$$m(s)=-K\,\frac{1+T_1s}{1+T_2s}\,\frac{V}{s} \qquad (54)$$

where

$$K=\frac{K_1K_2C_2}{C_1+C_2K_1^2}; \qquad T_1=\frac{T}{1+\sqrt{1+\dfrac{K_1^2C_2}{C_1}}}$$

$$\frac{V}{s}=v_1(s); \qquad T_2=\frac{T}{\sqrt{1+\dfrac{K_1^2C_2}{C_1}}}$$

Inverse transformation of equation 54 yields the minimizing function in the time domain:

$$m(t)=-KV\left[\left(\frac{T_1}{T_2}-1\right)\epsilon^{-t/T_2}+1\right] \qquad (55)$$

where ϵ is the base of the natural logarithms.

The corresponding transform of the variable q is readily obtained (see Fig. 4):

$$q(s)=-[K_2N(s)v_1(s)+K_1G(s)m(s)]=-\left[\frac{K_1}{1+Ts}\right]\left[\frac{K_2}{K_1}v_1(s)+m(s)\right] \qquad (76)$$

References

1. PROCESS AUTOMATION. *Report no. 1*, Case Institute of Technology, Cleveland, Ohio, Sept. 1956, pp. 13–14.

2. A. S. Boksenbom, R. Hood. *NACA TR1068*, National Aeronautics and Space Administration, Washington, D. C., 1952.

3. ENGINEERING CYBERNETICS, (book), H. S. Tsien. McGraw-Hill Book Company, Inc., New York, N. Y., 1954, pp. 198–213.

4. OPTIMIZING CONTROL OF A CHEMICAL PROCESS, D. P. Eckman, I. Lefkowitz. *Control Engineering*, New York, N. Y., vol. 4, Sept. 1957.

5. ANALYTICAL DESIGN OF LINEAR FEEDBACK CONTROLS (book), G. C. Newton, L. A. Gould. John Wiley & Sons, Inc., New York, N. Y., 1957.

6. CALCULUS OF VARIATIONS (book), R. Weinstock. McGraw-Hill Book Company, Inc., 1952, pp. 57–61.

7. AUTOMATIC FEEDBACK CONTROL SYSTEM SYNTHESIS (book), J. G. Truxal. McGraw-Hill Book Company, Inc., 1955, pp. 479–80.

8. OPERATIONAL CALCULUS BASED ON THE TWO-SIDED LAPLACE TRANSFORM (book), B. Van Der Pol, H. Bremmer. Cambridge University Press, Cambridge, England, 1950, p. 113.

9. METHODS OF THEORETICAL PHYSICS, PART I (book), P. M. Morse, H. Feshbach. McGraw-Hill Book Company, Inc., 1953.

A Method for the Design of Holding Electromagnets

J. T. LUDWIG
ASSOCIATE MEMBER AIEE

THE PROBLEM discussed is the optimum design of cylindrical holding electromagnets consisting of a ferromagnetic core, with stationary and movable portions, interlinking a coil of wire. These electromagnets are required to hold a switch or contact assembly in a given position as long as power is applied and to release it when power is removed. The electromagnet is not required to pull in the contact assembly; this is done by other means. It is desirable for the design method to proceed from the specified mechanical and electrical characteristics directly to the physical dimensions of the magnet without the use of trial-and-error methods.

The principal books on the subject are by Roters,[1] and Peek and Wagar.[2] Both treat the general problem of tractive magnet design from the trial-and-error analysis viewpoint. Here, the design of a holding electromagnet will be treated directly as a synthesis problem. The fundamental equations are used as a theoretical foundation for a straightforward design method.

A typical engineering design problem involves several conflicting factors, and its solution is usually a compromise between these conflicting factors. The usual method of attack on this type of problem is a trial-and-error approach. Certain aspects of the problem are guessed, estimated, or assumed; other aspects are sometimes calculated from these. This first trial solution is then analyzed by mathematical computation or by model measurement. If it does not meet the specifications, one or more of the assumed aspects are adjusted and the process is repeated until the specifications are met.

In some cases, this trial-and-error method is the only method of attack; but it is often used as an inferior substitute for more direct methods. The trial-and-error method usually wastes more engineering time and still does not lead to a design in which technically unspecified factors such as cost, size, weight, etc., are optimized.

When materials having nonlinear characteristics are involved, the direct design methods often make use of derived material characteristics. These are mathematical combinations of the traditionally measured fundamental characteristics. This approach has been used to relate the magnetic characteristics of core materials to their use in the magnetic circuits of inductors.[3–5] It is used here in the study of holding electromagnets.

Analysis, as distinguished from synthesis, is a calculation of the performance of a device when its physical dimensions are known. Synthesis, on the other hand, is the process of design and calculation of the physical dimensions of the inductor when its required performance is known. Analysis generally gives a unique solution. For a given operating point, a given device has a certain performance. Synthesis does not usually give a unique solution because an insufficient number of properties are specified. A large number of designs may meet a typical set of requirements. However, if the set of requirements is sufficiently complete, this large number will be reduced. The number of suitable designs will usually be reduced to one by an additional requirement to synthesize the smallest, the lightest, or the cheapest device meeting the other requirements.

Both calculations are useful in the design process since analysis is usually used as a check on the synthetic design. In this discussion, analysis will be presented first, and then synthesis.

The Basic Equations

The basic magnetic circuit under consideration is shown in Fig. 1. The hole in the center is provided for parts of the complete device. It is assumed that the iron cross-sectional area is everywhere constant. This would ideally require hyperbolic sections on the ends, but practical deviations from the ideal case probably do not introduce much error. For the cylindrical portions:

$$a_i = \pi(r_2^2 - r_1^2) \tag{1}$$

$$a_i = \pi(r_4^2 - r_3^2) \tag{2}$$

where a_i is the iron area in square centimeters, and r's are the radii shown in Fig. 1.

It should be noted that a mixed system of units is being used. All the magnetic and geometric units are in the centimeter-gram-second system while the electrical units are in the meter-kilogram-second practical system. This mixed system of units is used since, in U.S. practice, nearly all electrical specifications are in meter-kilogram-second units and nearly all magnetic material data are in centimeter-gram-second electromagnetic units.

The effective iron length is given by

$$l_i = 2c_1 t - r_1 - r_2 + r_3 + r_4 \tag{3}$$

where

l_i = effective iron length

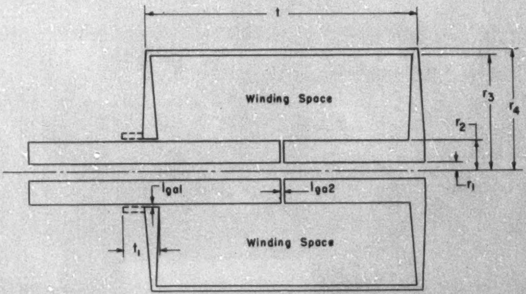


Fig. 1. Basic magnetic circuit

Paper 60-197, recommended by the AIEE Basic Sciences Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., January 31–February 5, 1960. Manuscript submitted October 29, 1959; made available for printing December 11, 1959.

J. T. Ludwig is with Electronic Communications, Inc., St. Petersburg, Fla.

The author wishes to acknowledge the help of Prof. N. Paul Stucky, Kansas Wesleyan University, Salina, Kans., and Dr. Ennis E. Jones, Minneapolis-Honeywell Regulator Company, Minneapolis, Minn., who processed the data in the production of the charts.

c_1 = dimensionless geometrical constant determined by the end section shape
l = physical length shown in Fig. 1 in centimeters

The total effective air gap length, ℓ_g, is the sum of the effective lengths of the annular gap around the plunger and the plane gap at the plunger end. The gap lengths considered here are only those in the closed position of the plunger, because the hold-in force depends only on the closed magnetic circuit conditions.

The effective length of the air gaps is shortened slightly by the fringing flux. The effective length of the annular gap may be shortened still more by providing greater magnetic area at the annular gap face. In both these situations, the effective gap length may be determined by calculating the gap reluctance and multiplying it by iron area, a_i. For the annular gap in Fig. 1, the effective gap length ℓ_{g1} is

$$\ell_{g1}=\ell_{ga1}\frac{a_i}{2r_2(\ell_{ga1}+l_1)} \tag{4}$$

where

ℓ_{g1} = effective length of gap 1
ℓ_{ga1} = actual length of gap 1
a_i = iron area in sq cm (square centimeters) of the core circuit, not the increased area of the air gap face, if any
r_2 = gap radius in cm
l_1 = gap width in cm

The plane plunger gap has an effective length of

$$\ell_{g2}=\ell_{ga2}\frac{a_i}{\pi\left[\left(r_2+\frac{\ell_{ga2}}{2}\right)^2-\left(r_1-\frac{\ell_{ga2}}{2}\right)^2\right]} \tag{5}$$

when

$$\frac{\ell_{ga2}}{2}\leqq r_1 \tag{6}$$

and

$$\ell_{g2}=\ell_{ga2}\frac{a_i}{\pi\left(r_2+\frac{\ell_{ga2}}{2}\right)^2} \tag{7}$$

when

$$r_1\leqq\frac{\ell_{ga2}}{2} \tag{8}$$

where

ℓ_{g2} = effective length of gap 2
ℓ_{ga2} = actual length of gap 2

These effective gap length equations are equivalent to assuming a band of fringing induction half a gap length wide outside the gap with the same induction as inside the gap. The total effective gap length is the sum of the individual gap lengths and is frequently conveniently expressed by a relative measure α defined by

$$\alpha=\frac{\ell_g}{l_i} \tag{9}$$

For the winding, the net copper area is

$$a_c=F_c c_2(r_3-r_2)l \tag{10}$$

where

a_c = net copper area in sq cm
F_c = space factor of the winding
c_2 = dimensionless geometrical constant determined by the end section shape

The mean length of turn (MLT) is

$$l_c=c_3\pi(r_2+r_3) \tag{11}$$

where

l_c = MLT in cm
c_3 = dimensionless geometrical constant determined by the end section shape

The outside surface area, through which the generated heat is transferred, is

$$a_s=2\pi(r_4^2+r_4 l) \tag{12}$$

where a_s is the surface area in sq cm

ELECTRICAL RELATIONS

If this core has a coil of N turns, the resistance of the mean turn is

$$R_t=\rho l_c N/a_c \tag{13}$$

where

R_t = resistance of the mean turn in ohms
ρ = resistivity of the winding conductor in ohm-cm
N = number of turns

The total copper resistance is then

$$R_c=\rho l_c N^2/a_c \tag{14}$$

where R_c is the total winding resistance in ohms.

If a direct voltage E is impressed on this coil, the steady-state current is given by Ohm's law,

$$I=E/R_c \tag{15}$$

where

I = current in amperes
E = direct voltage in volts

THERMAL RELATIONS

The heat generated in the winding is

$$W_c=I^2R_c \tag{16}$$

where W_c is the heat generated in watts. This heat must all be transferred through the outside surface so the heat power density is

$$K_c=W_c/a_s=I^2R_c/a_s \tag{17}$$

where K_c is the heat power density in watts per sq cm due to copper loss. Substituting in R_c from equation 14 gives

$$K_c=\frac{\rho l_c N^2 I^2}{a_c a_s} \tag{18}$$

MAGNETIC RELATIONS

The magnetic equations introduce another degree of complexity into the problem because of the nonlinear magnetic characteristics of the core material.

The force acting on the plunger in its closed position is

$$F=\frac{B^2a_i}{8\pi} \tag{19}$$

where

F = force in dynes
B = induction in gausses

This equation assumes negligible fringing of flux in the gap as would be the case when the core is closed and there is only a small gap between the plunger face and the stationary core face.

Also needed is the relationship giving induction B in terms of the current I and the physical dimensions of the electromagnet. The magnetic properties of materials are normally expressed in terms of the relationship between the induction B and the magnetizing force H.

In the air gap portions of the circuit B and H are equal. In the iron, the BH relationship is very complicated. It is neither linear nor single-valued. It depends on past history, and it is this complicated relationship that makes the design of magnetic circuits a difficult and esoteric art. Most magnetic characteristic curves represent some part of the

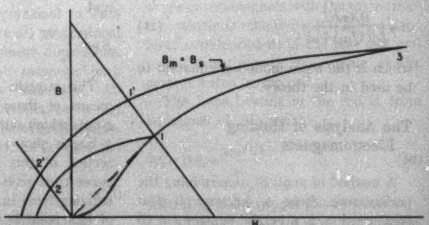


Fig. 2. Hysteresis loops and general air gap line analysis

BH function taken under particular conditions.

The most commonly available characteristic is the normal induction curve and, associated with it, the normal permeability. The normal induction curve is the locus of the tips of all symmetrical hysteresis loops, and Fig. 2 shows two of the symmetrical hysteresis loops that determine it. The co-ordinates of any point on the normal induction curve are then the maximum values of induction B_{mi} and magnetizing force H_{mi} for some symmetrical loop.

The normal permeability, μ_i, for any point on the normal induction curve is defined by

$$\mu_i = B_{mi}/H_{mi} \tag{20}$$

where

μ_i = normal permeability at the point in question

B_{mi} = induction in gausses at the point in question

H_{mi} = magnetizing force in oersteds at the point in question

This normal permeability is not a constant, but a variable; and it may be treated as a function of either B_{mi} or H_{mi}. Either the direct relationship of B_{mi} to H_{mi} or their ratio, μ_i, will be used in steps of the theory as convenient.

So far, only symmetrical (maximum and minimum values equal negatives of each other) magnetizations have been considered. When unsymmetrical magnetizations are allowed, another degree of freedom is added to the system. Ideally, data on a large group of minor loops would facilitate accurate analysis and synthesis. Often such complete data is not available; but some information can be gleaned from the fact that the tip of a minor loop farther from the origin lies on the normal induction curve.[6–9]

Referring to the magnetic circuit of Fig. 1, the cross-sectional iron area, a_i is essentially uniform around the circuit. Since the flux is continuous, the induction (flux density) is uniform around the circuit. For such a circuit

$$B = \frac{0.4\pi NI}{\ell_i[(1/\mu_i)+\alpha]} \tag{21}$$

which is the basic induction equation to be used in the theory.

The Analysis of Holding Electromagnets

A method of analysis, determining the performance from a known physical configuration, is a helpful supplement to the design procedure in providing a check on designs and in evaluating the effect of changes of the various physical dimensions.

The analysis problem is to answer the question: Given a device and its specified operating conditions, what is its performance or output?

In this case, the device is a holding electromagnet of known dimensions including core and coil. The operating conditions are the input voltage, current, and ambient temperature. The performance or output is the hold-in force when the coil is energized and the drop-out force due to residual induction which tends to hold the plunger in after the coil is de-energized.

The thermal aspects of the problem—calculating hot spot and mean temperature—are not detailed in this report, but reference is made to other literature on coil design.[10]

In analyzing a holding electromagnet, the first step is determining the current which flows through the coil. Depending on the external circuitry, the current, *I*, may or may not be influenced appreciably by the coil resistance. In the usual case of constant voltage power supply, the current is determined solely by the coil resistance, which must first be calculated for the operating temperatures by the usual methods. The current can then be obtained by Ohm's law, equation 15.

Karapetoff[11] described a method for finding the induction in a uniform magnetic circuit with air gaps when a known d-c magnetomotive force is applied. First, it is assumed that the relation between *B* and *H* in the iron is known and plotted as in Fig. 2. If the circuit is magnetized by a magnetomotive force, $0.4\pi NI$, the magnetomotive force equation is

$$0.4\pi NI = H\ell_i + B\ell_g \tag{22}$$

This is the equation of a straight line on the *BH* plane having intercepts

$$H = \frac{0.4\pi NI}{\ell_i} \tag{23}$$

and

$$B = \frac{0.4\pi NI}{\ell_g} \tag{24}$$

This straight line may be plotted by means of these intercepts or any two points which satisfy equation 22. Such a line is shown in Fig. 2. The intersection of this line with the iron *BH* curve then gives the induction and magnetizing force in the iron. The *B* value of this point is the induction and magnetizing force in the air gap.

To find the hold-in force, this method is applied to the normal induction curve, and the resulting induction *B* value is used in equation 19 to calculate the hold-in force.

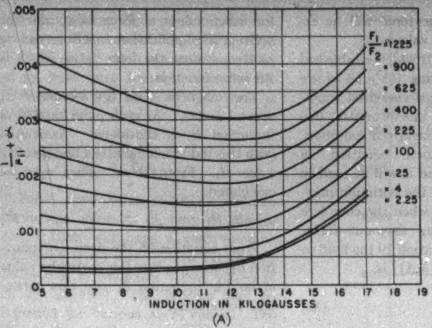
To find the drop-out force which tends to keep the plunger held in after the coil is de-energized, the appropriate descending branch of the *BH* characteristic must be used. Fig. 2 shows the Karapetoff line construction with the descending sides of two hysteresis loops as well. When the coil current is interrupted, the air-gap line moves to the left parallel to itself until it passes through the origin. Such a line is shown in Fig. 2. The induction is now that represented by the intersection of this second air-gap line with the appropriate descending *BH* curve. If the core had been magnetized to the value shown by the intersection with the first air-gap line at point 1, the de-energized induction would be that at point 2. One other possibility exists, however.

Sometime while the coil was energized, a voltage increase drove the air-gap line up to a higher value and the induction was that represented by point 3. When the voltage returned to normal, the operating point followed the saturated descending branch to point 1′. The induction is that represented by point 1′. When the coil is de-energized, the induction is that represented by 2′. If the latter value of induction is used in equation 19 to calculate drop-out force, the answer will be a maximum value of drop-out force. The actual operating value will always be equal to or less than this value since no possible magnetic history can arrive at an operating point above the descending branch of the saturated hysteresis loop. The de-energized operating point would lie on the line between 2 and 2′ for intermediate overvoltage occurrences.

It should also be noted that these drop-out induction values are only limits approached by the induction when the plunger remains closed. In an actual device, the restoring spring starts moving the plunger and opening the air gap when the induction force drops to the spring force value. On the air-gap line diagram, this is equivalent to rotating the air-gap line through the origin counterclockwise, which reduces the induction even lower.

Synthesis of Holding Electromagnets

The fundamental or traditionally measured characteristics of materials are not always in the most useful form for synthesis. This is particularly true when the



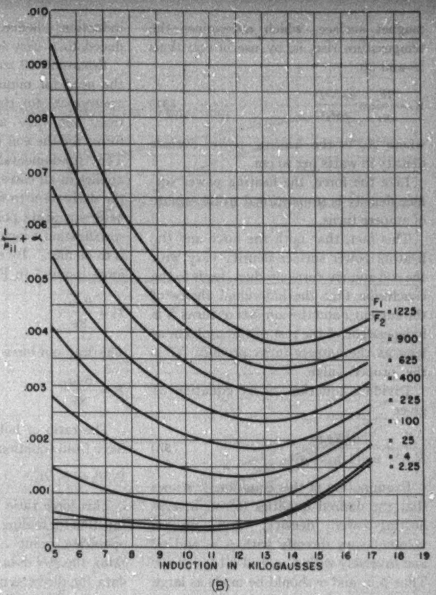


Fig. 3. Effective reluctivity curves for annealed Swedish charcoal iron

A—With no saturation (overvoltage) protection

B—With saturation (overvoltage) protection

characteristics are extremely nonlinear as in the case of ferromagnetics. An index function to guide the designer to the optimum operating point may be derived from the fundamental characteristics. Other application functions having unique values for each materials operating point may also be derived. If a graph is now constructed with the index function as the abscissa and the application functions as ordinates, the synthesis process becomes direct and straightforward. From the specifications, the index is calculated, the corresponding application functions are read from the graphs, and the physical dimensions are calculated.

This general approach requires that the index functions be matched only to specified quantities and not quantities which the designer sets out to determine. In other words, the synthesis process for the optimum holding electromagnet should not require the use of the core size in the calculations to find the core size.

The application functions are defined in a manner that makes the subsequent calculations as simple as possible.

When a holding electromagnet is to be designed, the following items will be known from the other considerations of the system:

Operating voltage
Ambient temperature
Allowable temperature rise
Duty cycle
Hold-in force
Drop-out force
Central hole size

The thermal problems may be treated separately to obtain a value of heating-power surface density through the outside surface of the magnet. This problem is discussed in other literature.[10]

In the mathematical formulation of the problem, it is possible and convenient to separate the determination of the optimum magnetic operating point and the determination of geometry (size and shape). The theory of the optimum magnetic operating point will be presented first and then the theory of the geometry will follow.

Optimum Magnetic Operating Point Effective Reluctivity: Optimum Size Criterion

The optimum magnetic operating point can be found by considering a holding electromagnet of fixed shape and finding the operating point which results in the smallest size. This operating point is independent of the shape. In the synthesis process, the optimum operating point is first determined, and then the size and shape are determined.

Consider the general geometry shown in Fig. 1. For a fixed shape, the linear dimensions are all proportional to each other, and the areas are all proportional to the squares of the linear dimensions. In this case, the mean magnetic path length is the reference dimension. All other lengths are proportional to it, and all surfaces are proportional to its square. If the mean iron length is ℓ_i, then the mean length of turn

$$\ell_c = u\ell_i \tag{25}$$

The net copper cross-sectional area,

$$a_c = v\ell_i^2 \tag{26}$$

The net iron cross-sectional area,

$$a_i = w\ell_i^2 \tag{27}$$

The total surface area for heat transmission,

$$a_s = p\ell_i^2 \tag{28}$$

The numbers, u, v, w, and p describe the shape while ℓ_i describes the size of the magnet. It should be noted that the number v includes the effect of coil space factor.

If the values of induction B from equation 21 and area a_i from equation 27 are inserted into force equation 19,

$$F = \frac{0.02\pi w N^2 I^2}{[(1/\mu_i)+\alpha]^2} \tag{29}$$

This equation shows that, for fixed shape electromagnets with the same effective magnetic reluctivity $[(1/\mu_i)+\alpha]$, the force is independent of the core size and varies directly with the square of the ampere turns in the coil.

The Joule heating in the coil is, from equations 14, 25, and 26,

$$W_c = I^2R_c = \frac{u\rho N^2 I^2}{v\ell_i} \tag{30}$$

where W_c is the copper loss in watts.

The density of heat flow through the

magnet surface, which determines the temperature rise, is, by use of equations 28 and 30,

$$K_c=\frac{W_c}{a_s}=\frac{u\rho N^2I^2}{pv\ell_i^3} \qquad (31)$$

where K_c is the heating power surface density in watts per sq cm.

Like the force, the heating power surface density is proportional to the square of ampere turns.

This fact, that both the force and the heating power surface density vary with the coil ampere-turns product, leads to the conclusion that the individual choice of the current I and the number of turns N is independent of the rest of the problem as long as the ampere-turns product NI is the proper value.

Dividing equation 29 by equation 31 gives

$$\frac{F}{K_c\ell_i^3}+\frac{0.02pvw}{u\rho[(1/\mu_i)+\alpha]^2} \qquad (32)$$

Examination of this equation 32 shows that the desired qualities (force, inverse heating-power density, and inverse volume) vary directly with p, v, and w; and inversely with u, ρ, and $[(1/\mu_i)+\alpha]^2$. Thus p, v, and w should be made as large as possible; u, ρ, and $[(1/\mu_i)+\alpha]$ should be made as small as possible. For a fixed shape electromagnet with a given surface area treatment, with or without fins, and with a given conductor material, the only quantity to be optimized is $[(1/\mu_i)+\alpha]$, the reluctivity, which is to be made as small as possible.

DETERMINATION OF EFFECTIVE RELUCTIVITY: GENERAL CASE

The quantity $[(1/\mu_i)+\alpha]$ requires detailed consideration. If the hold-in force (coil energized) were the only criterion, the best design would be that with minimum air gap α and the value of induction that gives maximum permeability μ_i. The hold-in force is not the only specification, however. The drop-out force when the coil is de-energized is also specified. It must not exceed some given value, in order that the return (spring) mechanism can function. The residual induction in the iron tends to maintain this force when the coil is de-energized. The air gap reduces this drop-out force.

The problem is then that of minimizing $[(1/\mu_i)+\alpha]$ subject to the constraint that the drop-out force (coil de-energized) be at or below some specified value. In actual operation, this value of drop-out force will not exist for more than an instant because the restoring spring will act before this force is reached, and the motion of the plunger will quickly reduce the induction; hence, the force will be reduced to a very low value.

Equations 29 and 32 which established the need for minimizing $[(1/\mu_i)+\alpha]$ are appropriate for the coil energized condition, but they are indeterminate 0/0 forms for the coil de-energized condition. The fundamental force equation 19 applies in all cases, and it will be used in the derivation to find the optimum magnetic operating point when the complete specification, including drop-out force, is to be met. By equation 19 the hold-in force [point 1 in Fig. 3(A)] is

$$F_1=\frac{B_1^2a_i}{8\pi} \qquad (33)$$

The drop-out force [point 2 in Fig. 3(A)] is

$$F_2=\frac{B_2^2a_i}{8\pi} \qquad (34)$$

The ratio of hold-in force to drop-out force from equations 33 and 34 is

$$F_1/F_2=(B_1/B_2)^2 \qquad (35)$$

This force ratio will become the index function for finding the optimum magnetic operating point. By use of the force ratio, the *BH* data can be translated into data for direct synthesis (design charts).

A fixed induction ratio derived from equation 35 is associated with a fixed force ratio,

$$B_1/B_2=\sqrt{F_1/F_2} \qquad (36)$$

The magnetic data can now be examined in terms of this fixed induction ratio, and the minimum value of $[(1/\mu_i)+\alpha]$ can be found for this value of induction ratio and the corresponding value of force ratio. The process can now be repeated for other values of force ratio, and the appropriate application functions can be plotted against the force ratio, which is the index function.

The magnetic data are examined at several points. At each point, a hold-in induction B_1 is considered. Associated with this induction value is a magnetizing force H_1. From these values may be calculated

$$\mu_{i1}=B_1/H_1 \qquad (37)$$

From equation 36 the drop-out induction B_2 corresponds to this particular value of hold-in induction B_1 and this particular force ratio,

$$B_2=B_1\sqrt{F_2/F_1} \qquad (38)$$

The demagnetizing leg of the hysteresis loop descending from point $(H_1,\ B_1)$ will at some point pass through the value of drop-out induction B_2 calculated by equation 38. If this occurs in the first quadrant, the particular combination of assumptions is not realizable. If, on the other hand, the appropriate demagnetizing leg of the curve passes through the calculated value of drop-out induction in the second quadrant, an air-gap line could be drawn through this point and the origin, and this point then represents the drop-out point of a magnetic circuit with relative air gap,

$$\alpha=-H_2/(B_1\sqrt{F_2/F_1}) \qquad (39)$$

where H_2 is the iron magnetizing force in oersteds corresponding to induction $B_2=B_1\sqrt{F_2/F_1}$ on the demagnetizing leg of the hysteresis loop descending from point $(H_1,\ B_1)$.

Since the point in question is in the

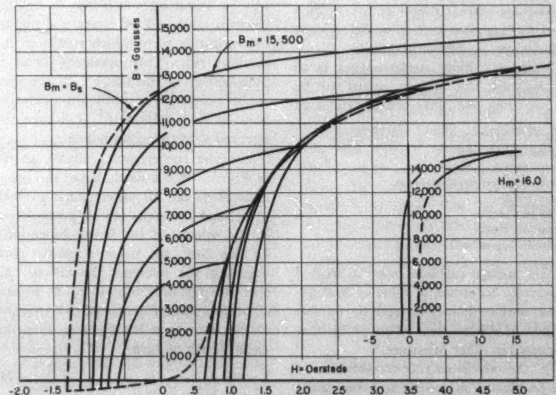


Fig. 4. Hysteresis loops for annealed Swedish charcoal iron

second quadrant, H_2 is negative and α is positive. For these assumed values of F_1/F_2 and B_1, the corresponding values of μ_{f1} have been found by equation 37. The quantity $[(1/\mu_{f1})+\alpha]$ may then be computed

$$\left(\frac{1}{\mu_{f1}}+\alpha\right)=\frac{H_1}{B_1}-\frac{H_2}{B_1}\sqrt{\frac{F_1}{F_2}} \qquad (40)$$

For a fixed value of F_1/F_2, a series of various values of B_1 may be considered. Each will have a corresponding value of $[(1/\mu_{f1})+\alpha]$. These values may be plotted as ordinates versus values of B_1 as abscissas. The process may be repeated for various F_1/F_2 values. The data in Roters' Electromagnetic Devices[12] are reproduced here as Fig. 4; the results of these calculations are shown in Fig. 3(A).

DETERMINATION OF EFFECTIVE RELUCTIVITY: PROTECTION AGAINST OVERVOLTAGE SATURATION

In a manner similar to that used for analyzing a holding electromagnet for dropout force under possible overvoltage saturation conditions, the magnetic data may be processed for synthesizing magnets having prescribed force ratios F_1/F_2, when subjected to overvoltage saturation conditions. Instead of using the demagnetizing portion of the hysteresis curve descending from the (H_1, B_1) point to locate the (H_2, B_2) point, the process uses the saturation demagnetization curve to locate the (H_2, B_2) point. This is the curve that the material would follow if it started at any unsaturated value and was subsequently saturated and then demagnetized. The results of these calculations are shown in Fig. 3(B).

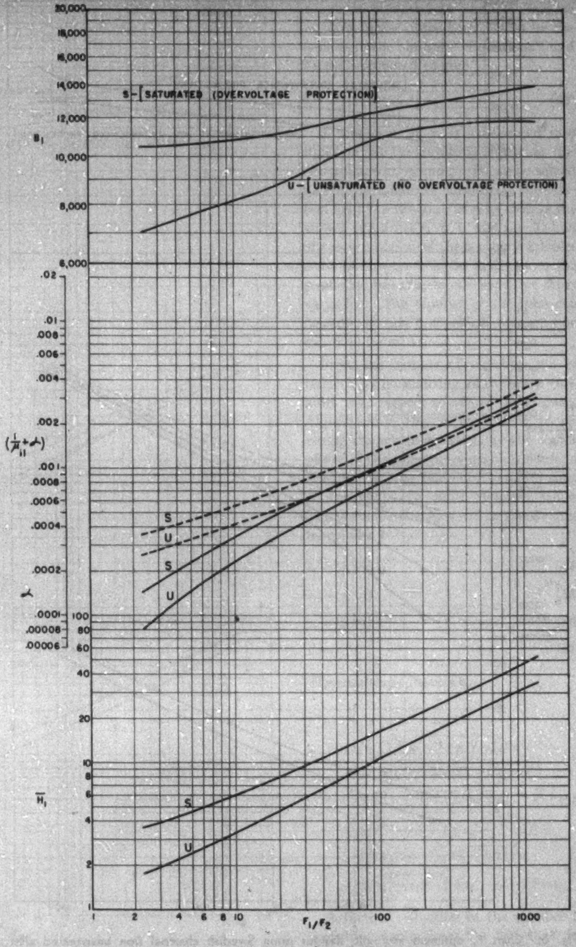


Fig. 5. **Chart 1, optimum magnetic designs using annealed Swedish charcoal iron**

CHART FOR OPTIMUM MAGNETIC OPERATING POINT

From the data calculated by the methods of the two previous paragraphs, and typified by Figs. 3(A) and (B), the optimum magnetic operating points are those of minimum $[(1/\mu_{f1})+\alpha]$. The data at the minimum points for the various materials and force ratios is most convenient for direct (free from trial-and-error methods) synthesis when plotted in the form of a design chart. Such a chart should have an index function for the abscissa and appropriate application functions for ordinates. In this case, the index function is the force ratio and the application functions are B_1, α, and a fictitious magnetizing force defined by

$$\overline{H}_1=B_1[(1/\mu_{f1})+\alpha] \qquad (41)$$

For each material there is a design chart with curves for designs with and without overvoltage saturation protection. Such charts are shown in Figs. 5 through 8. The curves labeled S (saturated) are curves in which saturation has been allowed and therefore represent self-protecting designs. For each chart, the data were taken from the minimum points of the corresponding effective reluctivity graph such as Fig. 3(A). For each force ratio point, the induction B_1 is obtained directly from the abscissa of the effective reluctivity graph. The value of α is obtained from equation 39 and the value of $\overline{H}_1$ from equation 41.

In use, the charts are entered at the specified index value, and the three application function values are read. These application function values are then inserted into the geometrical design procedures described later.

In addition, each chart has a comparison function $[(1/\mu_{f1})+\alpha]$ plotted as a dashed line for the cases with and without

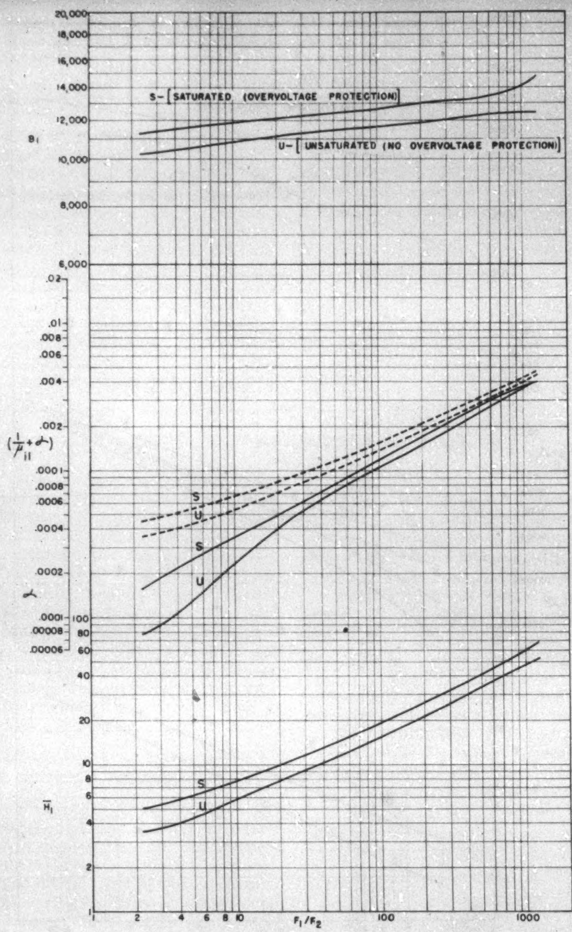


Fig. 6. Chart 1, optimum magnetic designs using Swedish charcoal iron unannealed after machining

saturation protection. The square of this function used in equation 32 will give the relative volumes occupied by magnets constructed of various materials.

AIR GAP LIMITED DESIGN

If the total effective air gap cannot be made as short as required for the optimum design with the specified F_1/F_2 ratio, a modified design procedure is necessary.

If the minimum physically realizable gap is very much larger than the one indicated by Chart 1, the optimum magnetic operating point is the point of maximum normal permeability, but the magnetic reluctance will then be nearly all in the gap, and there will be little advantage in operating on the precise maximum normal permeability point.

The maximum permeability point can readily be determined from the magnetic data of Fig. 2 by drawing a line through the origin tangent to the normal induction curve. The point of tangency is the point of maximum permeability. Such a line is shown dotted in Fig. 2. For this point, the permeability may be found from

$$\mu_{i1}=B_1/H_1 \tag{42}$$

Since the minimum physical gap is known, α can be estimated and $[(1/\mu_{i1})+\alpha]$ can be calculated. The application functions are now the known values of B_1, $[(1/\mu_{i1})+\alpha]$, α, and

$$\bar{H}_1=B_1[(1/\mu_{i1})+\alpha] \tag{43}$$

If the minimum physically realizable gap is only a little larger than the optimum indicated by Chart 1, the procedure of the last paragraph may result in too small an F_1/F_2 ratio. The optimum operating point will be farther above the point of maximum permeability and closer to the optimum point for gaps not having a limited minimum length. In this case, a design for the F_1/F_2 ratio corresponding to the minimum physically realizable α value will be very close to optimum.

OPTIMUM GEOMETRY

The optimum magnetic operating point determination is followed by the determination of the specific geometry. For some designs, it may be desired to realize absolute minimum volume. For other designs, either the length, t, or the outer radius r_4 may be specified, and the minimum value of the other (t or r_4) is desired. The method and chart here presented facilitate the solution of any of these problems.

From the specified hold-in force, F_1, and the optimum induction, B_1, the iron area, a_i, may be found from equation 19 as

$$a_i=8\pi F_1/B_1^2 \tag{44}$$

If a hole of radius r_1 is now specified in the center of the magnetic systems, Fig. 1, the plunger radius r_2 is then

$$r_2=\sqrt{r_1^2+\frac{a_i}{\pi}} \tag{45}$$

or substituting for a_i from equation 44

$$r_2=\sqrt{r_1^2+\frac{8F_1}{B_1^2}} \tag{46}$$

These last two equations include the case of no central hole by simply letting r_1 be zero.

If the value of a_c from equation 10, ℓ_c from equation 11, and a_s from equation 12 are inserted into equation 18, thus,

$$K_c=\frac{\rho c_3(r_2+r_3)N^2I^2}{2F_c c_2(r_3-r_2)(r_4+t)r_4 t} \tag{47}$$

From equations 22 and 43,

$$NI=(\bar{H}_1\ell_i)/(0.4\pi) \tag{48}$$

Inserting equation 3 into equation 48

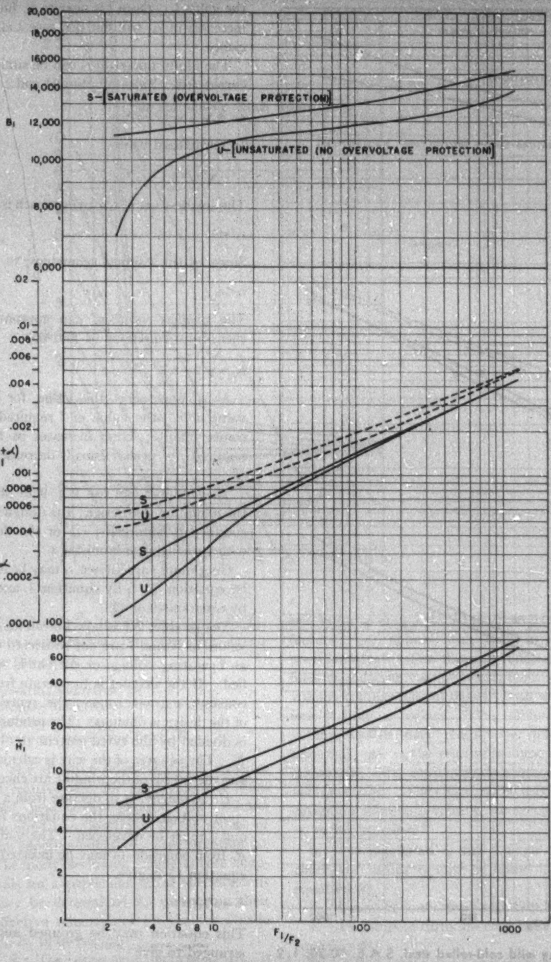


Fig. 7. Chart 1, optimum magnetic designs using mild cold-rolled steel, S.A.E. 10-10, dead soft unannealed after machining

$$NI=\frac{\bar{H}_1}{0.4\pi}(2c_1t-r_1-r_2+r_3+r_4) \qquad (49)$$

and equation 49 into equation 47 gives

$$K_c=\frac{\rho c_3\bar{H}_1^2(r_2+r_3)(2c_1t-r_1-r_2+r_3+r_4)^2}{0.32\pi^2F_cc_2(r_3-r_2)(r_4+t)r_4t} \qquad (50)$$

For simplification of notation, let a characteristic length, t_o, be defined by

$$t_o=\frac{\rho c_3\bar{H}_1^2}{0.08\pi^2c_2F_cK_c} \qquad (51)$$

Equation 50 may now be rewritten,

$$\begin{aligned}t^2[r_4(r_3-r_2)-c_1^2t_o(r_2+r_3)]+t[r_4^2(r_3-r_2)-\\ c_1t_o(r_2+r_3)(r_3+r_4-r_1-r_2)]+\\ (t_o/4)(r_2+r_3)(r_3+r_4-r_1-r_2)^2=0\end{aligned} \qquad (52)$$

This equation represents all the possible geometrical designs of the coil and outer core structure, the plunger radius having been determined by equation 46. In considering equation 52, the following quantities are known: r_1, r_2, t_o, and c_1. The unknowns are r_3, r_4, and t. Radius r_3 is not an independent variable because the area in the outer shell should be the same as in the plunger. Equations 1 and 2 may be equated to give

$$r_3=\sqrt{r_4^2-r_2^2+r_1^2} \qquad (53)$$

If this value of r_3 is substituted into equation 52 the resulting equation has two unknowns: r_4 and t. It can be solved for t by the quadratic formula. The solution for t is a function of r_4 as well as the known quantities r_1, r_2, t_o, and c_1. Tables of t as a function of these other variables could be prepared for use in designing the geometry, but they would have five other variables. The number of variables can be reduced, and a graphical presentation can then be made.

First, the number of variables can be reduced by normalizing all the other variables with respect to one of the knowns. Algebraically, this is equivalent to dividing the whole equation by the fourth power of the reference variable. Here, r_2 is chosen as the reference variable since it always has a finite nonzero value. The normalized values will be represented by primes:

$$r_1'=(r_1/r_2) \qquad (54)$$

$$r_4'=(r_4/r_2) \qquad (55)$$

$$t_o'=(t_o/r_2) \qquad (56)$$

$$t'=(t/r_2) \qquad (57)$$

The resulting equation is

$$\begin{aligned}t'^2[r_4'(\sqrt{r_4'^2+r_1'^2-1}-1)-\\ c_1^2t_o'(1+\sqrt{r_4'^2+r_1'^2-1})]+\\ t'[r_4'^2(\sqrt{r_4'^2+r_1'^2-1}-1)-\\ c_1t_o'(1+\sqrt{r_4'^2+r_1'^2-1})\times\\ (\sqrt{r_4'^2+r_1'^2-1}+r_4'-r_1'-1)]+\\ (t_o'/4)(1+\sqrt{r_4'^2+r_1'^2-1})\times\\ (\sqrt{r_4'^2+r_1'^2-1}+r_4'-r_1'-1)^2=0\end{aligned} \qquad (58)$$

A further reduction in the number of variables results from the fact that, for practical cases, t_o' is very small. Within the bracketed coefficients of t'^2 and t', the first terms predominate. The equation then becomes

$$\begin{aligned}t'^2[r_4'(\sqrt{r_4'^2+r_1'^2-1}-1)]+\\ t'[r_4'^2(\sqrt{r_4'^2+r_1'^2-1}-1)]-\\ (t_o'/4)(1+\sqrt{r_4'^2+r_1'^2-1})\times\\ (\sqrt{r_4'^2+r_1'^2-1}+r_4'-r_1'-1)^2=0\end{aligned} \qquad (59)$$

The positive root of equation 57 is

$$t'\cong\frac{t_o'(1+\sqrt{r_4'^2+r_1'^2-1})\times(\sqrt{r_4'^2+r_1'^2-1}+r_4'-r_1'-1)^2}{4r_4'^2(\sqrt{r_4'^2+r_1'^2-1}-1)} \qquad (60)$$

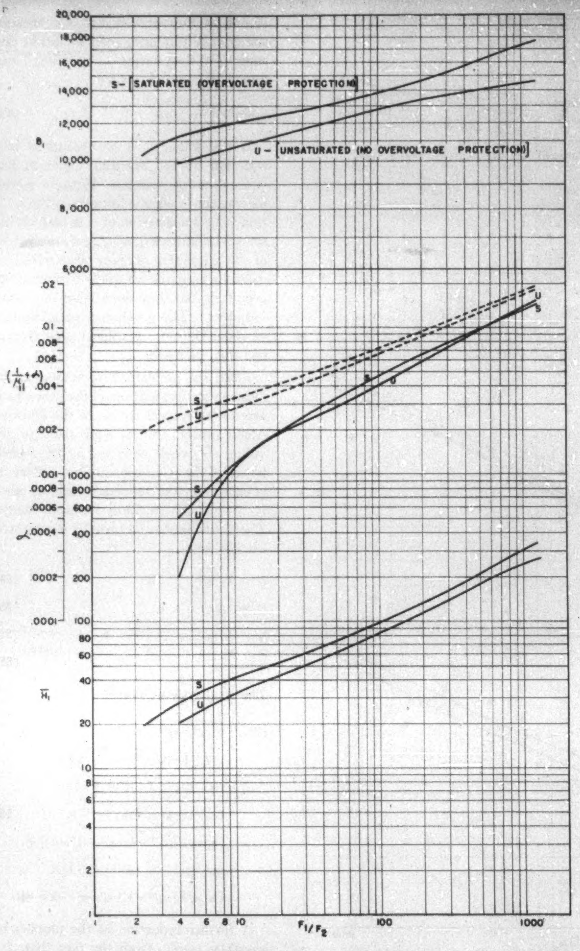


Fig. 8. Chart 1, optimum magnetic designs using mild cold-rolled steel, S.A.E. 10-20, 1/2 hard unannealed after machining

A general graph can be made by considering the ratio t'/t_o' which also equals t/t_o Thus,

$$\frac{t}{t_o}=\frac{t'}{t_o'}=\frac{(1+\sqrt{r_4'^2+r_1'^2-1})\times(\sqrt{r_4'^2+r_1'^2-1}+r_4'-r_1'-1)^2}{4r_4'^2(\sqrt{r_4'^2+r_1'^2-1}-1)} \tag{61}$$

Graphs of t/t_c versus r_4' for various values of r_1' are plotted in Fig. 9. The volume of the magnetic circuit is proportional to $r_4'^2\,t/t_o$. The diagonal lines are lines of constant $r_4'^2t/t_o$, and therefore constant volume lines. The relative volumes of various designs, resulting from various proportions of r_4 and t, may be determined directly from these lines.

The asymptotes of the curves are also of some value. For large r_4',

$$\frac{t}{t_o}\rightarrow 1 \tag{62}$$

In physical terms, no matter how much radial space is allowed for the circuit (unlimited r_4'), the length t will approach the value t_o. Once t is near to t_o, further increases in r_4' will not decrease t appreciably.

The useful ranges of r_4 can be similarly considered. From equations 1 and 2,

$$r_4.=\sqrt{r_3^2+r_2^2-r_1^2} \tag{63}$$

In normalized values,

$$r_4'-\sqrt{r_3'^2+1-r_1'^2} \tag{64}$$

The smallest value r_3' can approach is

$$r_3'=1 \tag{65}$$

Since, by the assumed geometry,

$$r_2<r_3 \tag{66}$$

The smallest value r_4' can approach is then, from equations 64 and 65,

$$r_4'=\sqrt{2-r_1'^2} \tag{67}$$

As r_4' approaches this value, for each value of r_1', the value of t required becomes infinite; large increases in t are required to permit small decreases in r_4'.

The designer can use Fig. 9 to obtain the minimum size design; the design with specified t, minimizing r_4; or the design with specified r_4, minimizing t.

Once r_1 and t are known, r_3 may be found by equation 53, ℓ_t by equation 3, and NI by equation 48 or 49.

The product NI is now known and the values of N and I are not restricted until an operating voltage or current is specified. If the magnet is to operate from a constant current source, the remainder of the design is obvious. The product NI is divided by the rated current to obtain N. The net area of the wire is calculated, and the details of the winding are checked.

If the magnet is to operate from a constant voltage source, the equations for N and I must be developed. The value of R_c from equation 14 may be inserted into equation 15 to give

$$E=\rho\ell_cN^2I/a_c \tag{68}$$

This equation may be grouped and rearranged to give

$$N=Ea_c/[\rho\ell_c(NI)] \tag{69}$$

and

$$I=\rho\ell_c(NI)^2/Ea_c \tag{70}$$

From equations 70 and 15,

$$R_c=E^2a_c/[\rho\ell_c(NI)^2] \tag{71}$$

SUMMARY OF SYNTHESIS STEPS

1. The first step in the synthesis of a holding electromagnet is the determination of the optimum magnetic operating point from Chart 1, Figs. 5 through 8. The index function (abscissa) is the ratio

Vol.
Pt. 1
196

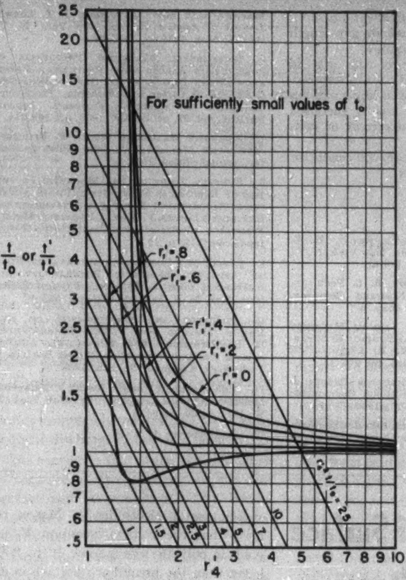


Fig. 9. Chart 2, geometrical designs

of hold-in force to drop-out force, F_1/F_2. Four application functions are given on the chart:

Operating induction, B_1
Effective reluctivity $[(1/\mu_{\Delta})+\alpha]$
Functional air gap, α
Effective magnetizing force, $\mathcal{H}_1$

The effective reluctivity is a measure of relative size, for fixed values of force and heat-power density, the volume required is proportional to the square of the effective reluctivity; see equation 32. The merit of various magnetic core materials for a particular set of specifications may be determined by comparing their effective reluctivities for the force ratio F_1/F_2, in question.

2. The operating induction, B_1, is used with the hold-in force, F_1, and the central hole size, radius r_1, to determine plunger size, radius r_2, by equation 46.

3. The effective magnetizing force, $\mathcal{H}_1$, is used, with the other known factors to calculate the characteristic thickness, t_o, by equation 51. This characteristic thickness is the asymptotic value that the thickness t approaches for large r_4 values.

4. Chart 2, Fig. 9 is a plot of the t/t_o versus r_4/r_2 with r_1/r_2 as a parameter. The thickness, t, has been normalized in terms of t_o; and the hole radius, r_1, and the outside radius, r_4, have been normalized in terms of r_2. Both the normalizing factors t_o and r_2 have been found in the previous steps. The diagonal lines are lines of constant volume which facilitate volume comparisons of designs with various t and r_4 values. The minimum volume design may readily be determined as well as the relative volume increase with shifts away from the minimum volume design. The chart can be used to design for minimum volume, for specified r_4, or for specified t.

5. Having found r_4 and t from the chart, the radius r_3 may be found from equation 53.

6. ℓ_i may be found from equation 3.

7. The ampere turns are computed by equation 48.

8. If the magnet is to operate from a constant current source, the rest of the design is straightforward. The number of turns can be found from the current and the ampere-turns, and the specific winding design worked out.

If the magnet is to be operated from constant voltage source, equations 69, 70, and 71 give turns, current, and resistance, respectively.

Conclusions

The theory developed presents direct methods for analysis and synthesis. These methods are free from successive approximation or trial-and-error procedures.

The first type of chart, Chart 1, provides a direct evaluation of the relative merits of various ferromagnetic materials. If, for example, for a given force ratio, F_1/F_2, two materials have reluctivities in the ratio of 2:1, the holding electromagnets would have volumes in the ratio 4:1 (two squared).

The second type of chart, Chart 2, shows the various possible geometrical proportions and sizes after the magnetic operating point has been established from the first type of chart.

The method is, of course, limited by the available magnetic material data. The minor loop (incremental) data is not always readily available.

Nomenclature

a_c = net (copper) winding area in sq cm
a_i = net iron area in sq cm
a_s = total outside surface area in sq cm
B = general induction in gausses
B_1 = induction in gausses when the coil is energized with normal design current or voltage
B_2 = induction in gausses when the coil is de-energized without allowing plunger movement
B_g = induction in the gap in gausses
B_{mi} = maximum induction in the iron in gausses
c_1 = dimensionless iron length geometrical constant determined by the end section shape; see equation 3
c_2 = dimensionless copper area geometrical constant depending on end section shape; see equation 10
c_3 = dimensionless mean-length-of-turn geometrical constant depending on end section shape; see equation 11
E = applied voltage in volts
F = general force in dynes
F_1 = hold-in force in dynes
F_2 = drop-out force in dynes
F_c = space factor of the (copper) winding
H = general magnetizing force in oersteds
H_1 = magnetizing force in oersteds in the iron when the coil is energized
H_2 = magnetizing force in oersteds in the iron when the coil is de-energized without allowing plunger movement
H_g = magnetizing force in the gap in oersteds
H_i = general magnetizing force in the iron in oersteds
H_{mi} = maximum magnetizing force in the iron in oersteds
$\mathcal{H}_1$ = fictitious effective magnetizing force in oersteds, $0.4\,\pi NI/\ell_i$
I = general current in amperes
K_c = copper loss heating power density in watts per sq cm
ℓ = general length in cm
ℓ_c = mean length of turn (copper) in cm
ℓ_g = total effective (plunger closed) air gap length in cm
ℓ_{g1} = effective length of gap 1 (annular) in cm
ℓ_{g2} = effective length of gap 2 (plane) in cm
ℓ_{ga1} = actual length of gap 1 (annular) in cm
ℓ_{ga2} = actual length of gap 2 (plane) in cm

ℓ_i = effective iron length in cm
M = general magnetomotive force in gilberts
N = number of turns
p = dimensionless geometrical ratio a_s/ℓ_i^2
R_c = (copper) winding resistance in ohms
R_t = resistance of the mean turn in ohms
r_1 = central hole radius in cm
r_1' = dimensionless central hole radius ratio r_1/r_2
r_2 = plunger outside radius in cm
r_3 = outer shell inside radius in centimeters
r_3' = dimensionless, outer shell inside radius ratio r_3/r_2
r_4 = outer shell outside radius in cm
r_4' = dimensionless, outer shell outside radius ratio r_4/r_2
t = outer shell length in cm
t' = dimensionless, outer shell length ratio t/r_2
t_0 = characteristic thickness (outer shell length) in cm, $\rho c_3 \bar{H}_1^2/0.08\,\pi^2 c_2 F_c K_c$
t_0' = dimensionless, characteristic thickness (outer shell length) ratio t_0/r_2
t_1 = width of gap 1 (annular) in cm
u = dimensionless geometrical ratio ℓ_c/ℓ_i
v = dimensionless geometrical ratio a_c/ℓ_i
W_c = heat power generated by the (copper) winding in watts
w = dimensionless geometrical ratio a_i/ℓ_i^2
α = dimensionless fractional air gap length ℓ_g/ℓ_i
μ_i = general permeability of the iron in gausses per oersted
μ_n = permeability of the iron in gausses per oersted when the coil is energized with normal design current or voltage
ρ = resistivity of the (copper) coil conductor in ohm-centimeters

References

1. Electromagnetic Devices (book), **H. C. Roters.** John Wiley & Sons, Inc., New York, N. Y., 1941, chap. X.

2. Switching Relay Design, **R. L. Peek, Jr., H. N. Wagar.** D. Van Nostrand Company, Inc., Princeton, N. J., 1955, chap. 9.

3. Optimum Air Gap for Various Magnetic Materials in Cores of Coils Subject to Superposed Direct Current, **V. E. Legg.** *AIEE Transactions*, vol. 64, Oct. 1945, pp. 709–12.

4. A. New Method for Inductor Design, **J. T. Ludwig.** *Proceedings*, Electronic Components Conference, Los Angeles, Calif., 1955.

5. Design of Optimum Inductors Using Magnetically Hard Ferrites in Combination with Magnetically Soft Materials, **J. T. Ludwig.** *Journal of Applied Physics*, New York, N. Y., vol. 29, Mar. 1958, pp. 497–99.

6. Incremental Magnetic Properties of Silicon-Iron Alloys with Particular Reference to the Design of Air-Gapped Smoothing Chokes, **R. O. Carter, D. L. Richards.** *Proceedings*, Institution of Electrical Engineers, London, England, vol. 97, pt. II, Apr. 1950, pp. 199–214.

7. Magnetic Materials (book), **F. Brailsford.** "Methuen's Monographs," John Wiley & Sons, Inc., second edition, 1951.

8. Further Study on the Magnetic Properties of Electrical Sheet Steel Under Superposed Alternating Field and Unsymmetrical Hysteresis Losses, **Y. Niwa, J. Sugiura, J. Matura.** *Researches*, Electrotechnical Laboratory, Tokyo, Japan, no. 144, p. 19.

9. Effect of a Superposed Alternating Field on Apparent Magnetic Permeability and Hysteresis Loss, **T. Spooner.** *Physical Review*, New York, N. Y., vol. 25, 1925, pp. 527–40.

10. Research and Development of New Design Method for Power Transformers, **H. L. Garbarino.** *ARF Project no. E-017-C*, Armour Research Foundation, Illinois Institute of Technology, Chicago, Ill., Feb. 1953.

11. The Magnetic Circuit (book), **V. Karapetoff.** McGraw-Hill Book Company, Inc., New York, N. Y., 1911, p. 29, prob. 11.

12. See reference 1, p. 50.

The Measurement and Influence of Surface Charges on the Inception Voltage Between Dielectric and Metal Dielectric Surfaces

S. I. REYNOLDS
NONMEMBER AIEE

THE CHARGES left on the surface of solid insulating materials which have been subjected to an electrical discharge were first discovered by Lichtenberg when he observed the peculiar formation of resin dust on the surface of a resin sheet which had been exposed to discharges.[1]

Many investigators since that time have studied charged surfaces by dusting techniques, by photographic plates exposed to the discharge, and by heat development of certain materials. They have shown by these methods that the extension and shape of the patterns are determined to some extent by the magnitude and polarity of the voltage transient. In addition semiquantitative measurements have been made by induced charge techniques with a suitable electrometer.

In this report results of measurements carried out by means of the rotating probe electrometer are given which show the magnitude and polarity of charges existing on small areas of a dielectric after breakdown has occurred between dielectric and metal dielectric surfaces.[2] Such results would be difficult to obtain using the other described methods.

With this new technique, charged areas as small as 0.3 cm (centimeters)^2 in area and 1 mm (millimeter) apart can be measured at a rate fast enough to show changes within intervals of 30 milliseconds. The minimum detectable charge density is about 0.6×10^{-9} coulomb cm^{-2}.

Finally, the methods used for charge measurements at the exact discharge inception voltage of a gap, the influence of these charges on the next breakdown, and its dependence on the polarity of the system are given.

Theoretical

The electric strength of a gas enclosed between dielectric surfaces is approximately the same as between equally spaced metal electrodes.[3,4] Mason reports that the voltage inception V_i for voids in polyethylene is some 10 to 20% lower than the breakdown voltage in air between metal electrodes, and refers to previous reports which give even lower values for cellulose acetate and photographic plate surfaces.

It is believed that the shape of the positive and negative "Lichtenberg" figures left on the surface of dielectrics, when the applied potential is of sufficient magnitude to produce ionization, is due to the greater mobility of electrons as they traverse paths in the gas adjacent to the dielectric surface.[5] This brings about a deposition or removal of electrons from the surface which depends on the polarity of the applied voltage. When the polarity is negative the bulk of the surface charge is supplied by emission from the metal cathode; when positive, electrons are removed from the dielectric.

Once a dielectric surface becomes charged in this manner, the charges transferred at the initial polarity are trapped. When the polarity is reversed the vector sum of the applied field and the reverse field due to these bound charges increases the field near the trapped charge and causes breakdown at a lower value of V_i.[6]

Mason has shown that if the initial

Paper 60-15, recommended by the AIEE Electrical Insulation Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., January 31–February 5, 1960. Manuscript submitted June 5, 1959; made available for printing October 7, 1959.

S. I. Reynolds is with the General Electric Company, Schenectady, N. Y.

Vol
Pt.
196

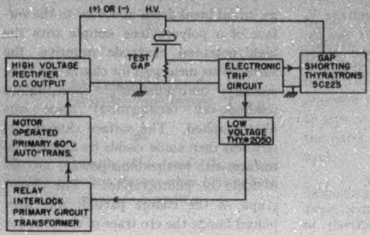

(+) OR (−)
H.V.
TEST GAP
HIGH VOLTAGE RECTIFIER D.C. OUTPUT
ELECTRONIC TRIP CIRCUIT
GAP SHORTING THYRATRONS 5C22'S
MOTOR OPERATED PRIMARY 60∿ AUTO-TRANS.
LOW VOLTAGE THY#2050
RELAY INTERLOCK PRIMARY CIRCUIT TRANSFORMER

Fig. 1 (left). Block diagram of high-voltage test set and gap

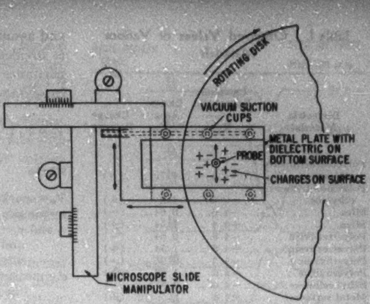

ROTATING DISK
VACUUM SUCTION CUPS
METAL PLATE WITH DIELECTRIC ON BOTTOM SURFACE
PROBE
CHARGES ON SURFACE
MICROSCOPE SLIDE MANIPULATOR

Fig. 3 (right). Rotating probe scanning device

breakdown occurs just above the inception voltage V_1 at one polarity, and if the polarity of the field is reversed, the breakdown voltage can then be reduced to values as low as 0.5 V_1.[6]

Thomas has shown that the charge left by an electric discharge of one polarity is not completely neutralized by a subsequent discharge of opposite polarity.[5]

He concludes that when the discharge is negative, electrons penetrate into the surface and become so strongly bound that they cannot be easily reached by positive ions.

Gross shows that the "homo" charge on electrets arises from discharges between electrodes and dielectric surfaces, and states that when breakdowns occur at one point the charge distribution of the dielectric at any other point remains unaffected if the surface charge is immobile.[7]

Purpose of Measurements

The foregoing theoretical discussion on breakdown voltage of a gas and its influence on surface charges produced on the surface of dielectrics as reported by a number of investigators, suggests that the following experiments may lead to a better understanding of these phenomena.

Fig. 2. Test gap with dielectric surface plate

With a slowly rising direct voltage applied, the discharge inception voltage of a uniform field gap will be determined by short-circuiting the gap by means of an electronic circuit, initiated by the first discharge. A series of tests will be made to compare the inception voltage of a metal to metal gap to that of a metal to dielectric gap using a number of different dielectric surfaces.

It will be shown that, when the breakdown occurred very close to the inception voltage, the area of surface charge remaining on the dielectric surface is smaller than the test electrode area and within the uniform field. The polarity, position, and magnitude of the surface charge will be determined by means of the rotating probe electrometer. Following the d-c tests, a-c breakdowns will be made just above the inception voltage for one and one-half cycles, between dielectric surfaces to determine the polarity of charge on the opposite and adjacent dielectric surfaces in its relation to the polarity of the last one-half voltage cycle.

Measurement of Surface Charge

The magnitude and polarity of small charged areas formed on dielectric surfaces by a discharge can be measured by means of the rotating probe electrometer.[2] The measurement technique is described under Apparatus and Procedure.

The voltage induced on the probe as it passes over the surface at a distance of 1 mm is capacitively coupled to the input of a vacuum tube electrometer, then amplified and presented on an oscilloscope where it can be calibrated and photographed.

From measurements of the input capacity to the electrometer, C (with the probe connected), and the oscilloscope deflection for a measured peak to peak sine wave voltage applied across this input capacity, the charge, Q, on the probe leading to an oscilloscope deflection d can be calculated. This is given by $Q=Cd/k$, where $k=d/V_s$, V_s being the peak to peak sine voltage applied to C. If fringing is neglected, the charge density on the surface being scanned is given by $\sigma=Cd/Ak$, where A is the area of the probe (0.284×10^{-2} cm^2), and C is 189 $\mu\mu$f (micromicrofarad), and $k=0.125\times10^{-4}$ cm/volt scope deflection.[2]

To measure the area on which the charge resides, marker pips are generated by placing two small wire electrodes embedded in polystyrene 3.48 mm apart and located in such a manner that the probe passes beneath the wires between passages over the areas being measured. By this means the polarity of the battery connected to the wires determines the polarity of the charge, and the distance between pips is used to measure the charge diameter.

Apparatus and Procedure

An electronic circuit was developed in order to prevent the slowly rising direct voltage from exceeding the discharge inception voltage. The first discharge across the gap is used to trigger a hydrogen thyratron which short circuits the gap in parallel with it within a fraction of a microsecond. The remaining parts of the circuit open the primary of the supply transformer, and operate a motor driven autotransformer to reduce the voltage to zero. A block diagram of the complete test circuit is shown in Fig. 1.

In order to maintain a uniform field gap for breakdown tests and to manipulate the dielectric sheets for surface charge measurements, the insulating material is fastened to the surface of flat 5×6 cm dural plates 1 mm thick, by

Table I. Observed Values of Various Materials

Dielectric Material	Inception Voltage V_a Kv	Charge Area, Sq Cm	Charge Polarity
Polychlorotrifluoroethylene	−4.5	0.79	(−)
Polychlorotrifluoroethylene	+4.65	4.2	(+)
Mica	−4.6	0.64	(−)
Mica	+4.4	1.54	(+)
Polyester resin	−4.6	0.64	(−)
Polyester resin	+4.6	1.54	(+)
Polyethylene	−4.5	0.95	(−)
Polyethylene	+4.6	1.86	(+)
Ethyl cellulose	−4.5	0.87	(−)
Metal surfaces	±4.6		(±)

means of double-coated pressure-sensitive tape. The insulation surface in this manner is maintained flat in the uniform field gap facing the high potential electrode as shown in Fig. 2, and also later placed on the slide manipulator by means of suction cups shown in Fig. 3 for scanning with the rotating probe. Charge location, polarity, and magnitude can then be obtained by movement of the dielectric surface over the probe as it passes beneath the surface.

By this procedure any detectable unwanted charge which might influence the applied field can be detected and removed before initial breakdown tests are made. After a discharge has occurred, the surface charge is then measured by photographing the trace which is presented on the oscilloscope screen synchronized with the rotating disk.

This technique of measurement has been found suitable for all of the dielectric materials tested, because the resulting bound charges remain fixed with no measurable change in area or magnitude during any of the tests reported in this paper.

Test Results

Discharges at the Inception Voltage

A large number of d-c breakdown tests were made between a 6.4 sq (square) cm circular stainless steel electrode and a number of different insulating materials at a gap distance of 1 mm with the high-voltage electrode negative and positive using the test gap shown in Fig. 2. The direct voltage was increased at a rate of 400 volts/sec (per second), and the total applied voltage was recorded when the electronic circuit short-circuited the gap and cut off the supply voltage.

The sparking potential of the gap was then calculated using the a-c potential division formula which takes into consideration the time constant of the system and assumes that the applied voltage V_0 is divided capacitively between the capacitance of the gap and the series capacitance of the dielectric as follows[8]

$$V_0 = V_a\left(\frac{\tau_1-\tau_2}{\tau\epsilon'}+1\right)$$

where

V_a = sparking potential of the gap
τ = air gap setting in mm
τ_1 and τ_2 = thickness of the dielectric sheet and the sticker film respectively, in mm
ϵ^1 = measured dielectric constant of the test sheet plus the sticker film

The observed values of the inception voltage V_a and the approximate charge area are listed in Table I.

Surface Charge Effect on Breakdown

In order to determine the influence of positive and negative surface charges on the breakdown of a dielectric gap caused by the resulting enhanced field in the charged region, a series of tests were run starting with uncharged dielectric surfaces by measuring the inception voltage before and after field polarity reversal.

An example of results which are typical of many such tests repeated on insulating surfaces is illustrated as follows. After an initial breakdown at 4.6 kv to the surface of a polyethylene sample with the high potential electrode negative, the surface was measured for charge position, polarity, and pulse height, and the cro (cathode-ray oscillograph) trace was photographed. The surface charge pattern was then made visible by dusting the surface with lycopodium powder so that it could be photographed. The photograph of the charge pattern was then placed beside the cro trace of the rotating probe, as shown in Fig. 4(A). The charged surface was then placed in the breakdown gap, and with the polarity reversed, the inception voltage occurred at (+) 2.3 kv.

The resulting charge pattern and voltage trace shown in Fig. 4(B) indicate that the discharge took place at the previous charged site and was responsible for reducing the inception voltage to 50% of the initial value, with the remaining surface charge still negative, showing that the positive discharge did not completely neutralize the initial negative charge. A second positive breakdown at +3.6 kv is shown in Fig. 4(C), with charge pattern extended beyond the original site with a very slight negative voltage trace remaining. After a third positive breakdown

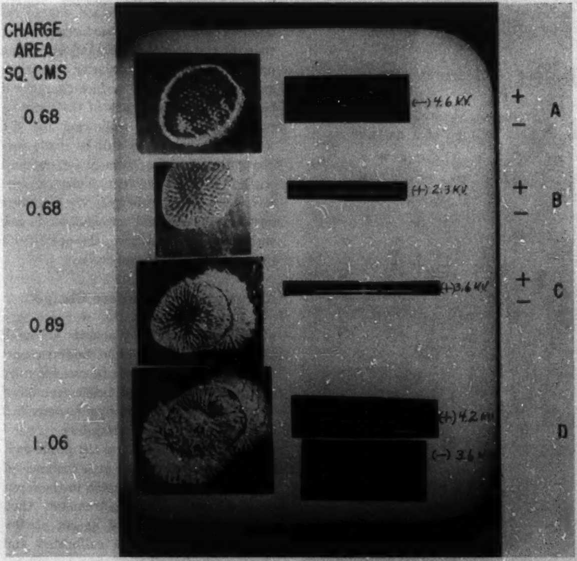


Fig. 4. Powder developed charges and cro trace of positive and negative discharge to polyethylene surface

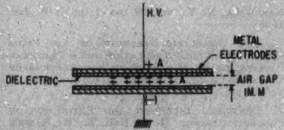


Fig. 5. Schematic of dielectric air gap and polarity of surface charges when high-voltage electrode is positive

at +4.2 kv, all of the negative charge appears to have been neutralized, but its effect on voltage breakdown still remains. This is shown Fig. 4(D) for two positions of the probe voltage trace across the charge pattern. Finally, a negative breakdown at −3.6 kv with charge probe voltages taken at three positions across the charge area are shown in Fig. 4(D). The charge pattern is the result of all the breakdowns made at this site.

CHARGE PATTERNS FROM ALTERNATING VOLTAGES

Results of the direct voltage tests show that the polarity of the charge on the dielectric surface opposite and facing the high-voltage electrode has the same polarity as the high-voltage electrode. If voltage breakdown tests are made with alternating voltages between parallel dielectric surfaces, the polarity on the surface opposite the high-voltage electrode should have the same sign as the last one-half cycle, and the charge on the other surface at the same site will be the unlike sign of the last half-cycle, as shown in Fig. 5.

Consequently, a series of experiments was run with 60-cycle voltages applied to dielectric surfaces separated by a 1-mm gap with the voltage just above the inception voltage for one and one-half cycles. The voltage was triggered on and off at zero voltage with an electronically controlled vacuum switch in series with the high potential winding of the supply transformer. The polarity of the last half-cycle was determined by an oscilloscope connected across a capacitor divider in parallel with the test gap.

Test results from breakdown tests between polyester resin surfaces are shown in Fig. 6, which shows the charge measured by means of the probe electrometer at several different positions on both dielectric surfaces.

Photographs of the oscilloscope trace numbered 1, 2, 3, and 4 show negative charges measured with the probe on the polyester resin surface opposite to the low voltage electrode which was negative on the last half cycle, and photographic traces of the polyester resin surface opposite the high-voltage electrode are shown in traces marked 5, 6, 7, and 8 located at the same positions and directly opposite to 1, 2, 3, and 4, respectively, to give positive charges in the surface with the same sign as the last half voltage cycle.

MEASUREMENT OF CHARGE MAGNITUDE FROM SINGLE DISCHARGES

Thomas has measured the total amount of charge left on a dielectric surface, using the lifted electrode technique of Gross. He found that the total charge from a negative breakdown at voltages above the inception voltage of a 0.5-mm gap and at rise times in the order of 0.4 microseconds using a 7.9-sq-cm uniform field brass electrode to be about 9,100 $\mu\mu$c (micromicrocurie), and total charge from positive wave front with same rise time to be about 6,200 $\mu\mu$c. He estimated that the central charged area had a negative charge of 7,000 $\mu\mu$c allowing 20% for the edge discharge.

By limiting the discharge from a metal electrode across a 1-mm gap to a dielectric surface in a uniform field, to the inception voltage V_i, we have found by virtue of the sharp resolution of the rotating probe that the negative and positive charge areas are much smaller than the test electrode area and that their location is within the central uniform field area.

This sharp resolution requires that the total charge must be measured by making many passes of the probe at intervals of 2 mm from one edge of the charged surface to the other in order to obtain a total integrated charge of the whole area.

A typical charge voltage trace corresponding to a given charge density is shown in Figs. 4 and 6. The height of the voltage trace on the oscillograph for a 1-cm deflection on the oscilloscope corresponds to a Ve value of 8 millivolts peak to peak of a sine wave across the input capacitance C of the electrometer. This corresponds to a Q value of 1.51×10^{-12} coulombs where $Q=CVe$.

Dividing this by the probe area of 0.284×10^{-2} cm^2 we obtain a charge density of 533×10^{-12} coulombs/sq cm for 1-cm cro deflection and calibrated at a probe distance of 1 mm. It has been determined that the width of the voltage trace at the points of inflection is within 8% of the true charge diameter. This width calibration is found by knowing the number of millimeters per division for the two marker pips which the probe also sees as it passes over the charged area. By scanning the charged area in one direction at intervals of 2 mm, starting at zero deflection on one side to zero deflection on the other side of the charged area, the charge Q for each sweep can be determined. Then, by adding the different charges for each sweep of the probe, the total charge for a measured area can be integrated.

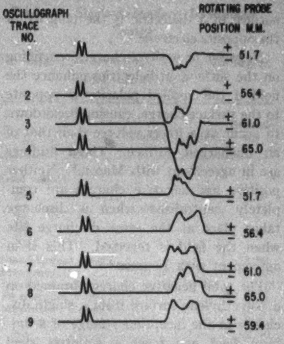


Fig. 6. Cro trace of charge patterns from one and one-half alternating voltage cycles

To determine the charge area for each sweep of the probe let x equal the voltage deflection in $\mu\mu$c/volt. Let y equal the charge width determined by the points of inflection as the probe passes over the charge area, then the $\mu\mu$c charge for each sweep is equal to $0.2\,xy$ cm, where 0.2 cm is the width of the probe slice. A good analogy of this measurement is to consider that the total charge corresponds to a loaf of bread which is being sliced by a bread slicer until the whole loaf is sliced. Each slice is weighed and the total weight corresponds to the total charge.

The total charge obtained from a negative and a positive charge remaining on a polyethylene surface obtained at the inception voltage of a 1-mm gap and measured with the rotating probe was measured to be 1,260 $\mu\mu$c for a +4.6-kv positive breakdown and 1,300 $\mu\mu$c for a negative breakdown at −4.7 kv.

Conclusions

The discharge inception voltage of a 1-mm gap in a uniform field between a metal electrode and a number of different insulating materials when the metal electrode is negative or positive has been found to be the same as the sparking potential between metal electrodes at 20 C (degrees centigrade) and 760-mm pressure in agreement with results obtained by Dubsky and Hall and Russek.[3,4]

The polarity of the charge left on the

surface of a dielectric is the polarity of the opposite electrode.

Negative or positive charges remaining on the surface of dielectrics enhance the field if the applied polarity is opposite to the surface charge, causing breakdown to occur at a lower voltage than that of an uncharged surface. These findings are in agreement with Mason.[6] Neither positive nor negative charges are completely neutralized when a discharge takes place at the same discharge site when the field is reversed. This is in agreement with Thomas.[5]

The total negative charge obtained on a polyethylene surface from a single discharge at the inception voltage of a uniform field gap from a stainless steel electrode is 1,300 $\mu\mu$c and, from a positive charge, 1,260 $\mu\mu$c. Thomas found a greater charge from a negative overvoltaged uniform field gap than from a positive breakdown.[5] Unlike Thomas it would appear that the negative and positive charges are essentially the same when the discharge is limited to the inception voltage in the uniform field.

It is hoped that this work can be continued with dielectric surfaces which have been subjected to ionization for extended periods of time and in the presence of different types of gases.

References

1. THE ATOMPHYSICAL INTERPRETATION OF LICHTENBERG FIGURES AND THEIR APPLICATION TO THE STUDY OF GAS DISCHARGE PHENOMENA, **F. H. Merrill, A. Von Hippel.** *Journal of Applied Physics*, New York, N. Y., vol 10, 1939, p. 873.

2. ROTATING PROBE ELECTROMETER, **J. C. Devins, S. I. Reynolds.** *Review of Scientific Instruments*, New York, N. Y., vol. 28, no. 1, pt 11-B, Jan. 1957.

3. THE DIELECTRIC STRENGTH OF AIR FILMS ENTRAPPED IN SOLID INSULATION AND A PRACTICAL APPLICATION OF THE PROBLEM FOR ALTERNATOR COILS AND CABLES, **B. Dubsky.** *AIEE Transactions*, vol. 38, pt. I, Feb. 1919, p. 537.

4. DISCHARGE INCEPTION AND EXTINCTION IN DIELECTRIC VOIDS, **H. C. Hall, R. M. Russek.** *Proceedings*, Institution of Electrical Engineers, London, England, vol. 101, 1954, pp. 47–55.

5. HEAT DEVELOPED AND POWDER LICHTENBERG FIGURES, AND THE IONIZATION OF DIELECTRIC SURFACES PRODUCED BY ELECTRICAL DISCHARGES, **A. M. Thomas.** *British Journal of Applied Physics*, London, England, vol. 2, 1951, p. 98.

6. BREAKDOWN OF INSULATION BY DISCHARGES, **J. H. Mason.** *Proceedings*, Institution of Electrical Engineers, London, England, vol. 100, pt. II-A, no. 3, 1953.

7. STATIC CHARGES ON DIELECTRICS, **B. Gross.** *Ibid.*, no. 1, 1950, p. 259.

8. FACTORS INFLUENCING THE DISCHARGE INCEPTION VOLTAGE IN GAS-SOLID DIELECTRIC SYSTEMS, **M. C. Halleck, J. C. Devins.** *Annual Report*, Conference on Electrical Insulation, National Academy of Sciences–National Research Council, Washington, D. C., 1954.

Electronic Transformer Design by Digital Computers

L. F. DEISE
NONMEMBER AIEE

WADE ETCHISON
NONMEMBER AIEE

REUBEN LEE
MEMBER AIEE

THE USE of computers in designing transformers for electric power transmission has been described previously.[1] For such transformers the design calculations are lengthy and may take many hours of engineering time. A large-scale computer requires only a few minutes to do the same work and is therefore of immense and obvious benefit. In contrast, small electronic transformer designs are simple, but a great many designs of these are used. To reap any advantage, computer design of electronic transformers must be rapid and easy, and adapted to the several kinds encountered in practice. This paper describes the procedures, results, and benefits pertaining to the application of digital computers to the design of electronic transformers.

Close co-operation between the engineer and the programer is necessary to achieve effective computer programs. The programer is usually a mathematician specializing in computer techniques. The precise method of storing information and using it differs with each computer and is so involved that it is a concern primarily of the programer and not of the engineer. In other words, if the engineer states his problem clearly, the programer can translate it into terms which the computer can use. This paper is concerned with the engineering use of computers in design work, rather than in the very involved mechanisms of computers. The main features of several computer programs are presented; more detailed descriptions are given in the appendixes and tables.

Simple 2-Winding Transformers

The starting point for the application of computers was a series of standard Hipersil cores which were tabulated, and maximum va (volt-ampere) rating assigned to each core based on experimental data, for the following conditions:

1. Line frequency of 60 cps (cycles per second).
2. Primary and secondary each single section.
3. Maximum voltage insulation in each winding, 1,000-volt operating value.
4. Resistive load.
5. Winding temperature rise 40 C (degrees centigrade) at ambient temperature of 65 C.
6. Maximum regulation 15%, no load to full load.
7. Open-coil assemblies encapsulated in polyester resin.
8. Equal copper loss in each winding.
9. Operating induction 16,000 gausses.

In adapting these simple transformers to computer design, provision is made for three primary voltages and four secondary voltages by means of winding taps, and for electrostatic shields between windings. At several points in the program, the computer is instructed to stop if individual calculations exceed practical limits. For example, the winding height, the temperature rise, or regulation or secondary terminal voltage limits might require the next larger core. On the other hand, if some of these quantities were below a minimum limit a smaller core might be used.

Usually the computer requires two iterations of programed design steps for the right answer. Sometimes it gives the right design the first time; sometimes it requires three iterations. The machine prints the following information for both primary and secondary windings:

1. Number of turns for whole winding and for each tap.
2. Wire sizes.
3. Number of layers.
4. Turns per layer.
5. Winding height.
6. Winding resistance.
7. Weight of wire.

Paper 59-120, recommended by the AIEE Electronics Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, February 1–6, 1959. Manuscript submitted July 10, 1958; made available for printing December 4, 1959.

L. F. DEISE, WADE ETCHISON, and REUBEN LEE are with the Westinghouse Electric Corporation, Baltimore, Md.

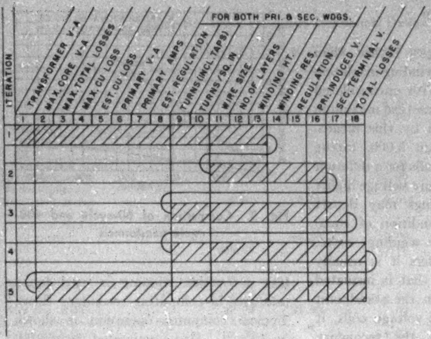


Fig. 1. Flow of work through computer for simple transformer design

8. Copper loss.

For the whole transformer the machine also prints:

1. Total losses.
2. Temperature rise.
3. Output voltages.
4. No-load to full-load regulation.

This program is shown in diagram form in Fig. 1. It consists of the following steps, each of which may have one or more substeps, adding up to a total of 87 programed calculations:

1. Transformer va rating is fed into the machine
2. A core is chosen with maximum va rating nearest to the transformer va
3. For the limiting temperature rise each core is assigned a maximum value for total losses
4. Maximum copper loss is step 3 minus core loss at assumed induction
5. Estimated copper loss = step 4 [(step 1)/(step 2)]^2
6. Primary va = step 1 + step 5 + core loss
7. Primary amp (amperes) = (step 6)/(primary volts)
8. Approximate regulation = (step 4)/(2.6 step 2) for each winding. (The factor 2.6 is 2×1.3, to change over-all regulation to regulation per winding, and to allow for maximum winding temperature; see Appendix I, step 29)
9. Primary turns are calculated for rated primary volts and an assumed value of core induction. Secondary turns = (secondary volts)/[(step 1+2δ) primary volts per turn], where

$$\delta=\frac{IR\text{ drop}}{\text{rated volts for either winding}}$$

where IR = winding resistance

10. Turns per square inch for each winding = ratio of turns to useful window area = N/WD_p
11. Wire sizes are picked from the wire table
12. Number of layers = turns/(turns per layer). Additional layers are required if taps are specified
13. Winding height = number of layers × (wire outside diameter + paper thickness)/(space factor)
14. Winding resistance is found from wire table, number of turns, and calculated mean length of turn
15. Regulation for resistance loads is sum of IR drops in percentage of rated voltages
16. Primary induced voltage is (primary terminal voltage)(1−δ)
17. Secondary terminal volts = (primary induced voltage × turns ratio)(1−δ)
18. Total losses = revised copper loss + core loss; should not exceed step 3

Note that copper loss in step 18 is calculated from actual currents and winding resistances, rather than the value estimated in step 5. Primary turns are based on rated primary volts in step 9. In Fig. 1 the action of the computer is shown for a design example having a va rating near, but slightly under, the maximum va rating assigned to a given core. Several winding taps are provided, and this increases the number of layers in each winding. Thus the winding height (step 13) is found to be too great for this core on iteration 1. The machine then repeats steps 11, 12, and 13 with the next smaller wire size. Now, the winding height is within limits, but the regulation is excessive. The machine then computes the primary induced voltage (step 16) and this allows the turns to be reduced in iteration 3. All limits are now met except secondary terminal volts, step 17. This step usually requires some trimming of turns in order to obtain the exact voltage. This is the final design, unless total losses exceed the allowed amount. If they should, the whole design would have to be done over on the next larger core, as indicated by iteration 5. Machine output for a simple transformer is given in Table I.

Core induction is chosen so that these transformers can be given the so-called overload test, in which voltage is raised

Table I. Single-Secondary Transformer Design

No Taps, Low-Voltage, 60 CPS, 40-Degree Rise, Open Type

	Primary	Shield	Secondary
Load voltage			
$E4$			12.00
$E3$	208.00		
$E2$			
$E1$			
$E0$			
I load current, amp			8.20
Frequency, cps	57.00		57.00
Insulation, inches	0.015	0.015	0.015
$E4$ tolerance, volts			0.12
Maximum regulation, volts/volt			0.1
$N4$			56.00
$N3$	889.00		
$N2$			
$N1$			
N/WD	1,764.918		111.176
Wire size, AWG	25.00		12.00
Layer insulation, 1 layer	1-002		3-005
Turns per layer	88.0		19.00
Number of layers			
$N1$			
$N1$ to $N2$			
$N2$ to $N3$	11.00		
$N3$ to $N4$			3.00
Total	11.00		3.00
Winding height, inches	0.239	0.005	0.282
Build-up per winding	0.269		0.269
Resistance, ohms	13.905		0.058
Wire weight, pounds	0.416		0.706
IR drop, volts	8.829		0.566
I^2R loss, watts	4.671		4.641
$E4$, computed volts			11.98
Regulation, volts/volt			0.094
Core			
Va	115.00		
Loss, watts	1.90		
Weight, pounds	1.90		
Area, square inches	0.89		
Allowance transformer loss, watts	13.30		
Width of winding, inches	1.876		
Winding margin, inches	0.156		
Coil form periphery, inches	4.938		
Width of coil, inches	2.188		
Temperature rise, C/watt	2.80		
Secondary, va	98.40		
Copper loss			
Allowable, watts	11.40		
Estimated, watts	9.754		
Primary, va	110.054		
Primary current, amperes	0.529		
Total winding height	0.526		
Build-up ratio	1.021		
Total winding build-up, inches	0.537		
Total I^2R loss, watts	10.057		
Total loss, watts	11.957		
Temperature rise, C.	33.48		
A-c flux density, gausses	15,412.919		

ol. 79
. 1
960

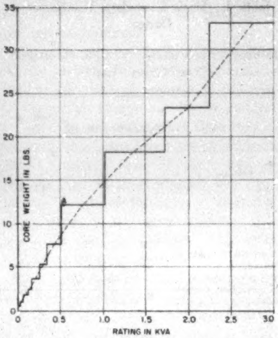


Fig. 2. Sixty-cycle transformer ratings versus core weight

12% above normal for duration of the test, without saturating the core. Coil margins, paper thickness, and coil form thickness are all fixed by winding-machine requirements. This minimum coil insulation with thorough encapsulation in a thermosetting resin, provides the necessary dielectric strength for 1,000 volts maximum.

Under these conditions, each core has the limiting va assigned to it on the program. Ratings increase stepwise, as shown in Fig. 2. Over a large number of designs, the average rating will be somewhere near the middle of the step as indicated by the irregular dashed line. Voltage regulation corresponding to the maximum rating in each case is shown in Fig. 3.

Essentially, these data represent experience; once they are incorporated in the computer program, the computer produces a maximum percentage of optimal, reliable designs. This percentage is never 100, because marginal ratings such as point *A*, Fig. 2, occasionally arise which might be designed better by a good engineer. But it is precisely to save the good engineer for more complex technical problems that the run-of-mine transformer designs are programed on the computer. Hence these cases are dealt with by one of three methods: 1. altering the program to accommodate the marginal case, 2. (if method 1 is not practicable, and size, weight or other attributes are important enough to warrant it) providing flags to detect the condition and checking for better results in longhand, or 3. (if the difference is trivial) by using the computer design, even if it deviates from optimum.

High-Voltage Program

For working voltages higher than 1,000, the margins and winding insulation must be increased. With encapsulated coils, 5,000 volts is the practical maximum which can be achieved by this means. At voltages higher than 5,000, taping must be used, but this calls for a different type of program. For any voltage higher than 1,000, the windings may depart materially from the condition of equal copper loss or equal winding space. In electronic transformers it is usually the secondary winding that is insulated for high voltage. From the standpoint of ease of winding high-voltage coils, it is preferable to place the secondary winding last. Of course, the core ratings for these high voltages are much less than Fig. 2 would indicate. Consequently, in this program, cores are assigned lower ratings as voltage increases.

Multiple Secondary Windings

A common instance of multiple secondary windings is the combined filament and plate transformer, usually requiring two or three secondaries. In setting up programs for this kind, it was found most convenient to proceed with the secondary turns calculations first, because for filament voltages the transformer has only a few secondary turns, and one turn is a large percentage of the total. Once the integral numbers of turns for these secondary voltages are established, then the primary turns may be calculated in reverse order. Regulation must be added to the secondary terminal voltage in computing induced voltage, and again in computing primary impressed voltage.

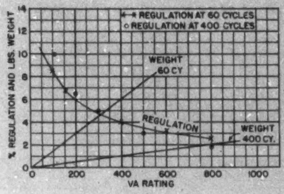


Fig. 3. Comparison of 60-cycle and 400-cycle transformers

Winding space is allotted to the various windings on the basis of turns and CM/amp (circular mils per ampere) which are given for each core. The last or outside winding is used for a reference in computing secondary turns. Typical computer operation is shown in Fig. 4. Since estimated regulation must be added to required output voltage to determine secondary induced voltage, frequently the calculated core induction on the second iteration is too high.

This requires an increase in secondary turns, and to obtain integral numbers of turns in all secondaries, the machine usually reduces core induction on the third iteration to a figure somewhat below the limiting value. Secondary turns are next computed inversely as the decrease in core induction and secondary voltages are found to be high. Then, on the fourth iteration, primary turns are changed accordingly. If equal CM/amp are used in all windings, the regulation problem is minimized, and all secondary voltages deviate from rated values by nearly the same percentage. This percentage is corrected on the fifth iteration.

If total losses are exceeded, the design process is repeated on the sixth iteration with a larger core. Fig. 4 shows the procedure in a condensed form. Calculation steps increase to 174. The total number of iterations increase as $n!$ where n is the number of secondary windings, but on a fast computer the

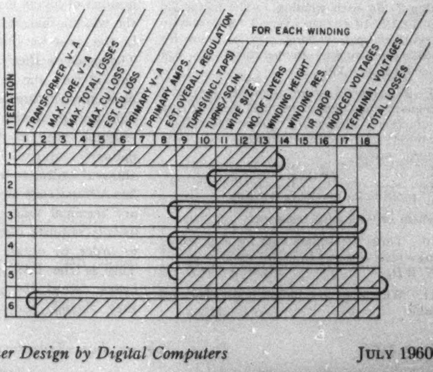


Fig. 4. Flow of work through computer for 3-secondary transformer design

Table II. Triple-Secondary Transformer Design

Low-Voltage, 60 Cps, 40 C Rise, Open Type

Load voltage					
*E*4			6.30	34.00	57.00
*E*3	220.00		3.15	17.00	28.50
*E*2					
*E*1					
*E*0					
I load current, amp			2.00	1.06	2.15
Frequency, cps	57.00		57.00	57.00	57.00
Va multiplier			1.00	0.707	0.707
Insulation, inches	0.015	0.015	0.015	0.015	0.015
*E*4 tolerance, volts					0.57
Maximum regulation				0.05	0.05
*N*4, turns		83.00	19.00	101.00	169.00
*N*3	626.00		10.00	51.00	85.00
*N*2					
*N*1					
Turns per volt					2.965
Wire, CM	556.603		1,800.00	954.00	1,935.00
Wire size, AWG	22.00	22.00	17.00	20.00	17.00
Layer insulation, one layer, mils	3		5	5	5
Turns per layer	83.00		48.00	65.00	48.00
Number of layers					
*N*1					
*N*1 to *N*2					
*N*2 to *N*3	8.00		1.00	1.00	2.00
*N*3 to *N*4			1.00	1.00	2.00
Total	8.00	1.00	2.00	2.00	4.00
Winding height, inches	0.243	0.03	0.105	0.078	0.211
Resistance, ohms	5.787		0.067	0.78	0.72
Wire weight, pounds	0.696		0.082	0.235	0.873
IR drop, volts	4.295		0.162	0.701	1.314
I^2R loss, watts	2.656		0.323	1.052	3.995
E_4, computed volts			6.385	34.101	56.92
Regulation, volts/volt			0.046	0.041	0.043
Core					
Va	260.00				
Loss, watts	3.90				
Weight, pounds	3.86				
Area, square inches	1.35				
Allowance transformer loss, watts	19.40				
Width of winding, inches	2.50				
Winding margin, inches	0.188				
Coil form periphery, inches	6.00				
Temperature rise, C per watt	2.10				
Total allowable build-up	0.782				
Secondary va	124.723				
Allowable copper loss, watts	15.50				
Estimated copper loss, watts	7.435				
Primary va	136.059				
Primary current, amp	0.618				
Total winding height	0.668				
Build-up ratio	1.058				
Voltage error indicator ratio	0.998				
Total I^2R loss, watts	8.668				
Total loss, watts	12.568				
Temperature rise, C	26.393				
A-c flux density, gausses	15,628.00				

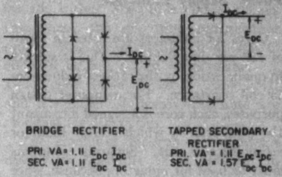

~
I_{DC}
+
E_{DC}
−
~
I_{DC}
+
E_{DC}
−
BRIDGE RECTIFIER
PRI. VA = 1.11 E_{DC} I_{DC}
SEC. VA = 1.11 E_{DC} I_{DC}
TAPPED SECONDARY RECTIFIER
PRI. VA = 1.11 E_{DC} I_{DC}
SEC. VA = 1.57 E_{DC} I_{DC}

Fig. 5. Single-phase rectifier circuits

increase in machine time is a matter of seconds. Printed output for a 4-winding transformer with center-tapped secondaries is given in Table II.

A single-secondary design run through this procedure shows that practically the same result is obtained by this method and that described in the section "Simple 2-Winding Transformers." Hence it is expected that the single-secondary program will be gradually replaced. Adding more than one secondary, of course, reduces the rating for a given core. It is feasible to modify this program for high voltages only if the high-voltage winding is located on top of the low-voltage windings.

Operating Frequencies

The foregoing programs are all for 60-cycle applications. For 400 cycles, or other frequencies, different core materials, ratings, operating induction, maximum temperature, and distribution of losses obtain. Hence it is necessary to modify the programs for these different frequencies.

As a practical matter, a double set of programs, one for 60 cycles and one for 400, has been sufficient for the large majority of applications to date. From the computer standpoint no additional problems are encountered in the 400-cycle programs. As indicated by Fig. 3, 400-cycle transformers weigh only a fraction of their 60-cycle counterparts, but have nearly the same regulation for a given rating.

Metal Enclosures

If for any reason a transformer must be enclosed in a metal case, the rating for a given core decreases, because of the inevitable temperature differential between the coil and the case. The ratio of the rating in a metal case to that in an open type of transformer for a given core may be less than 1 to 2. This ratio depends on the rating in several respects, and also on the kind of core material.

If it is possible to fill the metal case with punched laminations instead of type *C* cores, then laminations would be preferable. If a type *C* core is used, a double coil arrangement usually fills up the case to better advantage than a single coil. Each of these variations would require a special program of its own, and in order to avoid too many programs a reasonable compromise is to adopt one type of core material which covers most designs adequately. The kind of impregnating and filling materials also has a large influence on ratings.

Rectifier Plate Transformers

In the circuits of Fig. 5 are two common types of transformers used in single-phase rectifiers. In the case of bridge rectifiers, primary and secondary va are equal, so that the ratings are not appreciably different from those described in the section "Simple 2-Winding Transformers." In the tapped secondary case, the secondary va are appreciably greater than the primary, and the primary winding occupies less space than the secondary. Regulation loss and copper loss are figured differently for the secondary winding; hence for this kind of transformer two alternatives are open:

1. An engineer can interpret these secondary requirements and modify the resistive load program by reproportioning the winding space. This was done for the plate winding in Table II by means of the va multiplier.
2. A separate program can be made for rectifier transformers. In thyratron plate transformers, the transformer va ratings depart to a still larger extent from the output power, so that further modification is necessary.

Table III. Core Information for Transformer Design Program

12 Core Size	13 Nominal Core, Va	14 Core Loss, Watts	15 Core Weight, Pounds	16 Core Area, Square Inches	17 Allowable Transformer Losses, Watts	18 Width of Coil, Inches	19 Total Winding Build-up, Inches	20 Coil Form Periphery, Inches	21 Temperature Rise, C/Watt
1	4	0.2	0.19	0.178	0.9	1.063	0.403	2.50	7.4
2	8	0.35	0.34	0.267	1.85	1.313	0.347	2.688	5.2
3	25	0.6	0.60	0.356	4.4	1.875	0.507	3.352	3.6
4	50	1.1	1.06	0.595	7.1	1.875	0.484	4.375	3.2
5	115	1.9	1.90	0.890	13.3	2.188	0.597	4.938	2.8
6	175	2.7	2.63	0.980	16.8	2.875	0.822	5.25	2.4
7	260	3.9	3.86	1.35	19.4	2.875	0.822	6.00	2.1
8	340	5.5	5.44	1.84	22.4	2.875	0.822	6.75	1.8
9	510	7.8	7.80	2.14	27.0	3.75	1.16	7.313	1.5
10	1,120	12.2	12.2	2.91	42.5	4.063	1.384	8.844	1.0
11	1,720	18.3	18.3	3.43	57.0	5.50	1.892	10.00	0.7
12	2,250	23.5	23.5	4.41	65.5	5.313	1.722	11.125	0.6
13	3,230	33.1	33.1	4.23	82.5	8.5	2.34	10.25	0.5

Power Supply Inductors

Cores for inductors are rated in terms of LI^2 per unit volume. This is subject to qualification by the ratio of inductance to coil resistance. For computer design purposes, each core is assigned a constant $C=\sqrt{LI^2}\times L/R$, which relates inductor rating to core size. Separate programs must be used for inductors because, since these inductors carry direct current, both a-c and d-c components of induction must be calculated. In designing inductors by longhand, the designer may use these components in any proportion, so long as the sum does not exceed a limit fixed by the type of core material. In a computer program, turns are computed from the requirements of inductance I_{dc}, ripple voltage, and frequency. Then a-c induction, d-c induction, and air gap are calculated, and turns are adjusted, until the inductor is within rated limits.

Notes on Programs

Programs such as that of Appendix I are stored in the computer in binary form by means of either cards or magnetic tapes. To feed into the computer the data or specifications (operations 1 through 11 of Appendix I) for a new transformer design, an operator punches cards on a special machine. The cards are placed in the computer card reader, which then stores the data, also in binary form. As illustrated in Fig. 1, the stored program utilizes the data for the new design in performing the computations of steps 1 through 13 of Fig. 1. At this point, a comparison is made between the available winding height and the computed winding height for the wire used. If the computed winding height is the larger, the program automatically transfers to a section of the computer which reduces the wire size by one, and then returns to step 11; see Fig. 1. It will repeat this as many times as necessary to arrive at a computed winding height that is not greater than the available winding height.

At the end of step 16, Fig. 1, the primary induced voltage is checked in like manner with the specified voltage. If it is different by more than the permissible tolerance, the number of turns is adjusted and the program transfers back to the beginning of step 9. Similar checks are made on the secondary terminal voltage and allowable temperature rise. If the latter is too large, a new core is selected and the calculations are recommenced. All this is done automatically by the computer in accordance with the stored program. Other information stored in the computer will be mentioned in the course of the following program descriptions.

To refer to Appendix I, operations 1, 1a, 2, 3, etc., are items of information fed into the computer for a new transformer design. Voltages are arbitrarily given as zero volts for *E0P* and *E0S* and full rated voltage for *E3P* and *E4S*. Provision is for two primary taps *E1P* and *E2P*, and three secondary taps *E1S*, *E2S* and *E3S*, if taps are required. Voltages like *E3P* and *E1P*, terminal numbers, frequency, secondary current, voltage tolerances, and maximum regulation all change with the design requirements and represent input to the computer. For low-voltage designs (less than 1,000 volts on either winding) a minimum of 15 mils of paper insulation is provided by the program for primary, shield, and secondary for mechanical reasons. At 70 peak volts per mil this provides 1,000-volt insulation whether it is needed or not. Allowable temperature rise is the same for all transformers in this program and is 40 C. In operation 12 the core can be chosen by the computer from Table III (which is stored in another section). Associated with this core are several properties which have been determined by previous test or experience as follows (the operation numbers appear at the left):

13. Va rating.

14. Core loss. For a nominal flux density of 16,000 gausses, these cores have a limit on watts per pound of core loss, which, multiplied by the weight of the chosen core, gives a stored figure 14 for core loss.

15. Core weight.

16. Core area.

17. Allowable transformer loss (based on maximum rating 13).

18. Maximum coil width.

19. Total winding build-up.

20. Coil form periphery.

21. Temperature rise, C per watt loss.

22. Width of winding, calculated as indicated in the program. The margin is based on good winding practice, necessary to keep wire from falling into the margin and causing collapse of the wound coil.

23. Build-up per winding D_p is the allowable build-up for either primary or secondary, and contains allowances for insulation, coil form, shield, and clearances. Equal winding spaces are provided for primary and secondary. This does not provide exactly the same copper loss in both windings because ordinarily there is a difference of up to 20% in the average length of turn of the two windings. This difference in copper loss is at least partly mitigated by the fact that the outside winding is nearer to the ambient air, and consequently cools better.

All the information up to this point has been set up in the computer in tabular form, or has been fed in for a particular design. The computer goes to work with operation 24. Succeeding operations are self-explanatory except for the following:

29. The symbol δ represents the approximate regulation for each winding, from no load to full load, in volts per volt at room operating temperature. The factor 2.6 is 2×1.3. The factor 2 is to reduce from over-all regulation to regulation for each winding and 1.3 is the ratio of winding resistance for copper at 100 C to that at 25 C (the wire tables give winding resistance at 25 C). Operation 30 is an equation used when dimensions are in English units and flux density is in gausses. This is equation 34; see page 83 of reference 2. Turns *N3P* correspond to voltage *E3P*; turns *N4S* correspond to voltage *E4S*, etc. The symbol α is the property known as volts per turn.

Ratios such as 38 and 39 are really functions of the wire table. As used in these programs, they give turns per square inch of winding area $W\times D_p$ for each winding. It can be shown that N/WD_p=(turns per inch)/(wire outside diameter+paper thickness), or that the ratio N/WD_p is a function of the wire table. This ratio has been figured in Table IV (which is also stored in the computer) for wire sizes with cross-sectional areas from 10 to 36,500 CM. The column "Layer Insulation" of Table IV is the paper thickness used for wire sizes based on standard winding practice. From this table, the computer chooses the approximate wire size based on available winding space WD_p.

Operation 52 is a tabulation of standard winding practices to guide the winder. The practices include kind and length of leads, direction of windings, turn tolerances, and joint insulation.

In operations 54 through 56 and 60 through 63, a whole layer is used for the number of turns between taps even though it would be possible to get the required number of turns in a fraction of a layer. This is to allow taps to be brought out at the end of a layer, which is necessary on multiple-winding machines. However, if the taps are in the top layer, the layers may be reduced as in operation 64.

In operations 58 and 65 the revised winding build-ups are based on actual turns rather than on space available. If total build-up exceeds the available space, it is caught in operation 66, and wire sizes are then revised. If a wire shield (one layer of wire) has been specified, there may be insufficient room for it; this is also dealt with in operation 66, paragraph *c*.

Primary regulation 73 is figured at 75 C, which is an average operating condition; therefore copper resistances are multiplied by 1.2 in operations 73 and 78. However, to figure the total I^2R in step 80, an additional 1.08 is used to figure the maximum copper loss at 100 C. Operation 89 is a check on flux density.

Appendix I has been modified for several other programs. The first of these is for high-voltage (up to 5-kv) transformers. Where the original program specified the same winding insulation for all designs, the high-voltage program permits the engineer to specify definite thicknesses for primary, shield, and secondary insulation. Correspondingly, the winding margins are increased. These insulation thicknesses and margins affect the total winding build-up and coil width in operations 22 and 23. Increased space for the high-voltage secondary is provided in operation 24 by a proportionality factor. This affects the copper loss distribution, and a new table of core properties must be used in place of Table III.

Rectifier transformers of the kind shown at the right of Fig. 5 require larger space for the secondary than for the primary. This is handled by a "va multiplier," by which the computer may be instructed to redistribute the winding space so that the secondary has approximately 41% more winding area than the primary. For low-voltage transformers this does not affect operation 19, but does affect operations 24, 26 through 28, and subsequent winding calculations.

In the multiple secondary program, the computer provides a redistribution of winding space similar to that mentioned above, but based on the secondary va in each winding. Thus, the secondaries become the starting points for the program. This includes insulation as well as va rating. Initially, the turns are based on secondary rather than primary turns, and *IR* drops are added to the output and primary induced voltages, rather than subtracted from the primary impressed voltage.

Transformers of 400 cycles are designed on another set of cores wound from 4-mil rather than 13-mil strip. The program is similar to that of Appendix I, except that 80-degree temperature rise is allowed. Average ambient is figured at 20 C and maximum ambient at 70 C. These transformers are used in portable apparatus, where weight is especially important. They are characterized by small dimensions, light weight, and high operating temperature. With a large operating temperature range, regulation becomes important. No-load to full-load regulation is limited so that filament voltages may be maintained within the tube book values throughout the operating temperature range. Regulation is figured at 100 C and I^2R at still higher temperature. A 4-mil core table replaces Table III, and a number of design constants are modified for 400 cycles.

The computer program for inductor design in Appendix II is different in several respects from a transformer program. The inductors programed here are smoothing inductors for d-c power supply filters, with Hipersil cores, and are for maximum peak voltages of 1,000 or less. These are linear inductors, in which the air gap is large compared to l_c/μ, where l_c is the mean core length and μ is the core permeability. For such inductors, equations 35, 36, and 38 of reference 2 are valid. These equations can be combined to yield an expression for coil turns:

$$B_{dc}=0.6NI/l_g \text{ gauss} \tag{35}$$

$$B_{ac}=3.49E\times10^6/fA_cN \text{ gausses} \tag{36}$$

$$L=\frac{3.19N^2A_c\times10^{-8}}{l_g+l_c/\mu} \text{ henrys} \tag{38}$$

Adding equation 35 to equation 36,

Table IV. Wire Chart for Computer Programs

CM	Wire Size	Turns Per Inch	Winding Height Per Layer	$\frac{N}{WD_p}$	Resistance, Ohms Per 1,000 Feet	Weight, Pounds Per 1,000 Feet	Layer Insulation Thickness, Inches
36,500	0.081×0.365	2.5	0.099	24	0.289	113	0.005
28,600	0.102×0.229	3.7	0.119	31	0.369	89.6	
24,600	0.102×0.204	4.3	0.116	37	0.430	74.4	
20,000	0.091×0.182	4.6	0.106	43	0.529	62.6	
12,500	0.051×0.204	4.4	0.066	63	0.843	39.9	
10,700	0.064×0.144	5.8	0.078	74	0.990	33.8	
10,200	0.072×0.114	7	0.084	85	1.04	33.0	
8,070	0.081×0.081	9	0.093	95	1.312	25.4	
6,530	12*	10	0.0941	106	1.62	19.9	
5,178	13	12	0.0852	141	2.04	15.79	
4,107	14	13	0.0772	168	2.58	12.53	
3,257	15	15	0.0701	214	3.25	9.95	
2,583	16	17	0.0633	268	4.09	7.882	
2,048	17	19	0.0527	358	5.16	6.256	
1,624	18	21	0.0477	440	6.51	4.967	
1,288	19	23	0.0432	532	8.21	3.944	
1,022	20	26	0.0391	664	10.4	3.128	
810	21	30	0.0336	892	13.0	2.484	0.003
642	22	33	0.0304	1,085	16.4	1.973	
510	23	37	0.0274	1,350	20.8	1.564	
404	24	42	0.0249	1,685	26.2	1.243	0.002
320	25	47	0.0217	2,165	33.0	0.9876	
254	26	52	0.0195	2,660	41.6	0.7828	
202	27	57	0.0178	3,200	52.5	0.6221	
160	28	64	0.0157	4,070	66.2	0.4926	0.0015
127	29	71	0.0141	5,040	83.4	0.3921	
100	30	80	0.0128	6,250	105.2	0.3118	
79.7	31	88	0.0115	7,650	133	0.2471	
63.2	32	98	0.0106	9,250	167	0.1965	
50.1	33	110	0.0095	11,580	211	0.1552	
39.8	34	124	0.0082	15,120	266	0.1235	0.001
31.5	35	140	0.0074	18,900	336	0.0970	
25.0	36	155	0.0068	22,800	423	0.0779	
19.8	37	170	0.0062	27,400	533	0.0616	
15.7	38	193	0.0057	33,900	672	0.0491	0.00075
9.9	40	239	0.00445	53,700	1,069	0.0308	0.0005

* This and subsequent numbers in this column are AWG.

and neglecting l_c/μ in equation 38, yields

$$N=\frac{10^6(18.75LI+3.49E/f)}{(B_{dc}+B_{ac})A_c}$$

For a tentative design limit on $B_{dc}+B_{ac}$ of 15,000 gausses,

$$N=\frac{1}{A_c}\left(1{,}250LI+233\,\frac{E}{f}\right)$$

This relation is used in operation 20 of Appendix II.

For the inductor program, each core is assigned an inductor rating indicator $C=\sqrt{LI^2}\times L/R$. It can be shown that C is proportional to the fourth power of a core dimension (e.g., the core window width G), if the ratios of the same dimensions are maintained throughout the range of core sizes used. The computer is thus provided with a sensitive indicator of core size, and can select a core for any rating which is fed into it.

In Appendix II, computer operations 1 through 6 represent information fed in for a new inductor design. From this information the computer can calculate the size indicator C in operation 7 and thus select core 8. Associated with this core are properties 8 through 18 used in later computations.

There is but one winding in these inductors, so that build-up 14 is for the whole winding rather than part of it as in the transformer programs. Items 2a and 2b are tolerances on rated inductance. This is an important matter from the standpoint of manufacturing, because cores have variations in permeability, and air gaps are likely to vary from unit to unit, depending on the size of the gap. The rated inductance is the average of the two limits in operation 19. Inductor turns are based on this inductance in operation 20. The equation is based on a maximum flux density of 15,000 gausses. This is somewhat conservative for Hipersil cores but allows for adjustment to be made in operation 27. A fringing factor of 15% was included in the derivation. After the air gap is calculated in operation 26, fringing is indirectly refigured in operation 27 and the number of reactor turns adjusted accordingly.

Incremental permeability is stored in the computer as a separate double-entry table with two arguments: a-c and d-c induction. Intersections of rows and columns provide four values of incremental permeability which enclose the value to be used. The value used is then found by double linear interpolation in operation 30, with the use of operations 27 and 28.

Conclusions

Co-ordination of engineering and mathematicians' viewpoints is necessary before computer programs become effective. The International Business Machines Corporation *704* computer is presently used for these programs. Computing time averages one second, with an average of less than one minute for the operator, including machine input and output time. Algebraic functions are used most, which makes transformer programs easy to arrange. Such information as wire tables, core data, layer paper, and incremental permeability must be tabulated and stored in the computer.

Over-all gains are improvement in reliability because of better transformer standardization, great reduction in design engineering time, and more effective use of engineering talent. Because of the flexibility and adaptability of existing programs, it may be predicted that further applications of these programs to audio transformers, carrier, and pulse transformers will soon appear. Computer designs tend to be more uniform than those done by hand, and less dependent on individual whims. Therefore they are easier to check and control. It is possible to tabulate the computer output so as to produce manufacturing information without transcription, thus eliminating a chore distasteful to engineers. It is not difficult to envisage further transitions from the computer design of transformers to automatic manufacture and test.

Appendix I. Transformer Design Program

This transformer design program is concerned with a single secondary transformer of 60 cps, 40-degree temperature rise, and open-frame Fosterite coils under 1 kv.

The numbers at the left refer to computer operations in sequence. Heavy type indicates that the number is used in later operations.

1. $E3P$, volts...terminal no.
1a. $E0P$, volts...terminal no.
2. $E1P$, volts...terminal no.
3. $E2P$, volts...terminal no.
4. Frequency, cps
5. $E4S$, volts...terminal no.
5a. $E0S$, volts...terminal no.
6. $E1S$, volts...terminal no.
7. $E2S$, volts...terminal no.
8. $E3S$, volts...terminal no.
9. IS, amp
10. Secondary voltage tolerance, volts
11. Regulation maximum limit, volt/volt
12. Core type
13. Core va
14. Core loss, watts
15. Core weight, pounds
16. Core area, square inches
17. Allowable transformer losses, watts
18. Width of coil, inches
19. Total winding build-up, inches
20. Coil form periphery
21. Temperature rise, C per watt
22. Width of winding, inches $=\mathbf{18}-2\times$ margin $=W$
23. Build-up per winding $D_p=\dfrac{\mathbf{19}-3\times\text{insulation thickness}}{2}$
24. Va of secondary winding $=\mathbf{5}\times\mathbf{9}$
25. Allowable copper loss, watts $=\mathbf{17}-\mathbf{14}$
26. Estimated copper loss, $\mathbf{25}\times\mathbf{24}/\mathbf{13}$
27. Primary va $=\mathbf{24}+\mathbf{14}+\mathbf{26}$
28. Primary current $=\mathbf{27}/\mathbf{1}$ amp
29. $\delta=\mathbf{25}/(2.6\times\mathbf{13})$
30. $N3P=\dfrac{3.49\times\mathbf{1}\times10^6}{\mathbf{4}\times\mathbf{16}\times16{,}000}$ turns. (Use whole number; drop decimal.)
31. $\alpha=E3P/N3P=\mathbf{1}/\mathbf{30}$
32. $N4S=(1+2\times\mathbf{29})\mathbf{5}/\mathbf{31}$ turns. (Use whole number; drop decimal.)
33. $N1P=\mathbf{2}\times\mathbf{30}/\mathbf{1}$ turns. (Use whole number; drop decimal.)
34. $N2P=\mathbf{3}\times\mathbf{30}/\mathbf{1}$ turns. (Use whole number; drop decimal.)
35. $N1S=\mathbf{6}\times\mathbf{32}/\mathbf{5}$ turns. (Use whole number; drop decimal.)
36. $N2S=\mathbf{7}\times\mathbf{32}/\mathbf{5}$ turns. (Use whole number; drop decimal.)
37. $N3S=\mathbf{8}\times\mathbf{32}/\mathbf{5}$ turns. (Use whole number; drop decimal.)
38. $N3P/WDP=\mathbf{30}/(\mathbf{22}\times\mathbf{23})$
39. $N4S/WDP=\mathbf{32}/(\mathbf{22}\times\mathbf{23})$
40. Select primary wire size from Table IV by **38** (next number smaller than **38**)
41. Turns per inch for primary wire
42. Winding height per layer
43. Resistance in ohms per 1,000 feet
44. Weight in pounds per 1,000 feet
45. Primary layer insulation
46. Select secondary wire size from Table I by **39** (next number smaller than **38**); print wire size
47. Turns per inch for secondary wire
48. Winding height per layer
49. Resistance in ohms per 1,000 feet
50. Weight in pounds per 1,000 feet
51. Secondary layer insulation
52. Winding practices
53. Primary turns per layer $=\mathbf{22}\times\mathbf{41}$. (Use whole number; drop decimal.)

54. Number of layers for $N1P$=**33/53**. (Use next largest whole number.)

55. Number of layers from $N1P$ to $N2P$=(**34**−**33**)/**53**. (Use next largest whole number.)

56. Number of layers from $N2P$ to $N3P$=(**30**−**34**)/**53**. (Use next largest whole number.)

57. Number of primary layers=**54**+**55**+**56**

58. Height of primary winding=**57**×**42**

59. Secondary turns per layer=**22**×**47**. (Use whole number; drop decimal.)

60. Number of layers for $N1S$=**35/59**. (Use next largest whole number.)

61. Number of layers $N1S$ to $N2S$=(**36**−**35**)/**59**. (Use next largest whole number.)

62. Number of layers $N2S$ to $N3S$=(**37**−**32**)/**59**. (Use next largest whole number.)

63. Number of layers $N3S$ to $N4S$=(**32**−**37**)/**59**. (Use next largest whole number.)

64. Number of secondary layers=**60**+**61**+**62**+**63**. If **32**−**35** is less than **59**, subtract two layers from **64**. If **32**−**35** is more than **59** but **32**−**36** is less than **59**, subtract one layer from **64**. (Use revised value of **64** as determined above.)

65. Height of secondary=**64**×**48**. (Use revised value of **64** as determined above.)

66. Height of primary, secondary, and shield=**58**+**65**+**42**. This number must not exceed 2×**23**
    a. If the primary wire size is no. 22 or larger, and if 2×**23** is exceeded, check height of primary+secondary+solid shield as follows: **65**+**58**+0.005. If this is not right, go to the following condition:
    b. If **66** still exceeds 2×**23**, revert to **40** and choose primary wire and wire wound shield one size smaller. Test solid shield again if 2×**23** is still exceeded. If not right, go to the following conditions:
    c. If **66** still exceeds 2×**23** after adjustment 66b, revert to step **46** and choose secondary wire one size smaller. If it will not go with primary and wire wound shield, test with solid shield

67. Shield wire size: print AWG (American Wire Gage) size or solid as in operation **66**

68. Height of shield=**42** or 0.005 (based on **67**)

69. Build-up ratio=2×**23**/**66**

70. Primary mean turn=**20**+π **58** inches

71. Primary resistance = **30** × **70** × **43**/12,000 ohms

72. Primary wire weight=**30**×**70**×**44**/12,000 pounds

73. Primary IR drop=1.2×**28**×**71** volts

74. Primary I^2R loss=**73**×**28** watts

75. Secondary mean turn=**70**+π[**58**+**65**+2 (insulation+**68**)]

76. Secondary resistance = **32** × **75** × **49**/12,000 ohms

77. Secondary wire weight=**32**×**75**×**50**/12,000 pounds

78. Secondary IR drop=1.2×**9**×**76** volts

79. Secondary I^2R loss=1.2×**9**²×**76** watts

80. Total I^2R losses=1.08×(**74**+**79**) watts

81. Total losses=**80**+**14** watts

82. Total temperature rise=**21**×**81** C. This result should be less than 40 C. If it is not, choose core with next larger va rating

83. Primary induced voltage=**1**−**73** volts

84. Output voltage

$$E_s=\left(\frac{83}{30}\times 32\right)-78$$

If E_s is outside of tolerance given by **10**, adjust primary turns by

$$N3P=\frac{83\times 32}{5+78}$$

Then, revert to step **31** and repeat calculations

85. No-load output voltage=(**1**×**32**)/**30**

86. Regulation=(**85**−**84**)/**84**. If regulation exceeds that specified by **11**, choose next larger core size and repeat calculations

87. A-c flux density in gausses is

$$\frac{3.49\times 83\times 10^6}{4\times 16\times 30}$$

Appendix II. Computer Design of Linear Inductors

This computer design of linear inductors uses Hipersil cores, open-type Fosterite encapsulated coils up to 1.0 kv peak.

1. Maximum winding resistance, ohms
2. Inductance, henrys

2a. Minimum inductance multiplier

2b. Maximum inductance multiplier

3. Current, amperes
4. A-c voltage, volts (ripple)
5. Frequency, cps
6. Allowable temperature rise, C
7. C is

$$L/R\times\sqrt{LI^2}=\frac{2}{1}\times\sqrt{2\times 3^2}$$

8. Core type
9. Core weight, pounds
10. Core area, square inches
11. Temperature rise, C per watt
12. Width of coil, inches
13. Winding margin, inches
14. Winding buildup, inches
15. Coil form periphery, inches
16. Core mean length, inches
17. Core periphery, inches
18. Core G dimension, inches
19. Average multiplier=(**2a**+**2b**)/2
20. Number of reactor turns=

$$\frac{1}{10}\left(1{,}250\times 2\times 19\times 3+\frac{233\times 4}{5}\right)$$

(Use whole number only; drop decimal.)

21. A-c flux density, gausses=

$$\frac{3.49\times 4\times 10^6}{5\times 10\times 20}$$

22. D-c flux density, gausses=15,000−**21**
23. Total air gap, inches=(0.6×**19**×**3**)/**22**
24. Air gap per leg, inches=**23**/2. Round off to next largest number in third decimal place.
25. Effective gap per leg, in inches, is

$$24/\left(1+\frac{17\times 24}{2\times 10}\log_e\frac{18}{24}\right)$$

26. Number of reactor turns adjusted=

$$25\times\sqrt{\frac{25\times 2}{23}}$$

27. A-c flux density, in gausses adjusted=(**20**×**21**)/**26**
28. D-c flux density, in gausses adjusted=(0.5×**26**×**3**)/(2×**25**)
29. Total flux density, in gausses=**27**+**28**. This number must not exceed 16,000; if it should, choose next larger core size
30. Incremental permeability from table
31. Inductance in henrys:

$$\frac{3.2\times 26^2\times 10\times 10^{-8}}{2\times 25+(16/30)}$$

32. Minimum inductance limit=**2a**×**2**
33. Maximum inductance limit=**2b**×**2**
34. Item **31** must be more than **32** but less than **33**. If not within limits, multiply **26** by the factor$\sqrt{19\times 2/31}$ and repeat computer operations from **27** on. (Provide for two increases or decreases of turns)
35. $\dfrac{N}{WD_p}=\dfrac{26}{14(12-2\times 13)}$
36. Select wire size from Table IV by **35** (next number smaller than **35**)
37. Turns per inch for wire size
38. Winding height per layer
39. Resistance in ohms per 1,000 feet
40. Weight in pounds per 1,000 feet
41. Layer insulation as in Table IV. (Print thickness of layer insulation.)
42. Turns per layer=**37**×(**12**−2×**13**) (Use whole number; drop decimal.)
43. Number of layers=**26**/**42**
44. Height of winding=**38**×**43**. This number must not exceed **14**; if it should, use wire one size smaller.
45. Mean turn in inches=**15**+π**44**
46. Winding resistance in ohms=(**26**×**45**×**39**)/12,000. This number not to exceed **1**; if it should, use next larger core

47. Weight of wire in pounds $=(\mathbf{26}\times\mathbf{45}\times\mathbf{40})/12{,}000$

48. *IR* drop in volts $=\mathbf{3}\times\mathbf{46}\times1.2$

49. I^2R loss in watts $=\mathbf{3}\times\mathbf{48}$

50. Core loss in watts $=1.59\times10^{-11}\times\mathbf{5}^{1.89}\times\mathbf{27}^{1.84}\times\mathbf{9}$

51. Total reactor loss in watts $=\mathbf{49}+\mathbf{50}$

52. Total temperature rise in C $=\mathbf{11}\times\mathbf{51}$. This should not exceed **7**; if it does, use next larger core

53. Winding tolerance ±1%

References

1. Application of a Digital Computer to the Design of Power Transformers to Specification, **C. L. Moore, W. T. Duboc, P. A. Zaphyr**. *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 75, May 1956, pp. 134–38.

2. Electronic Transformers and Circuits, **Reuben Lee**. John Wiley & Sons, Inc., New York, N. Y., second edition, 1955.

◆

Discussion

L. W. Kirkwood (Bell Telephone Laboratories, Inc., Whippany, N. J.): The computer age is still in its infancy and a paper which indicates a computer application which offers a number of advantages is certainly valuable. It can provide the impetus for more general use of computers, especially in the field of small electronic transformer design. It should be recognized that this paper, first presented as a conference paper, was the first report involving the use of digital computers to design electronic transformers. The specific application here is electronic power transformers and of particular interest is an appreciation of the numbers of such design required. One would think that over the years a design would be available for about every circuit or system application that could be conceived, but it is only in isolated instances that a circuit or equipment designer will specify an existing design for a transformer supplier. It is true that many of these designs are "simple" variations of existing designs, such as other arrangements of windings on a given mechanical structure, but it is here that the computer program is a great asset. As the paper points out, a technician by simply punching a few appropriate holes in a card can relieve an engineer for more "glamorous" work. The design program contained in the Appendix is a significant contribution. There may be many programs considered to be proprietory, but the authors of the paper present transformer and inductor step-by-step programs in detail. A person not skilled in the art of power transformer design, and who may not have a digital computer, can easily follow the step-by-step program.

As stated, the program is flexible and can be readily adapted to the many varieties and kinds of transformers. It is very easy to adjust the program to emphasize any specific design characteristic, namely, regulation, size, weight, etc. The speed of operation is such that any number of complete designs can be obtained for as many numbers of variations one wishes to make. We have a program for an International Business Machines Corporation *704* computer that, once it is fed into the machine, will grind out a 4-secondary winding design in a few seconds. Because of the speed in obtaining an answer, the effects of any variation of any parameter on the over-all design can readily be determined simply by introducing changes on punch cards and rerunning the program for each change. Thus a program can be adjusted to do optimum designs in whichever area one wishes.

Cost and justification for a computer have been a matter of discussion. A computer for the type of design work discussed must be a large one. Aside from the storage of information required for many interrelated calculations, much space must be provided for the storage of complete wire tables, core loss and magnetizing current characteristics of cores, dielectric characteristics of insulation systems, heating data, etc. In some instances, where a *650* computer is not sufficient, one will require a *704* or its equivalent, but it is only in the larger companies doing a lot of work involving varied calculations, the sum total of which will justify their own machine.

Some, however, use the services of a neighbor. We did this temporarily, until our machine was installed. The growth of electronics and the increasing complexity of systems promotes a trend toward sharing effort. In some instances design programs are no longer considered proprietory. By extension, a standardized design program may be in a library of a computing center and can be rented, as is the case with many standardized services.

Allen B. Rosenstein and **Gerald Estrin** (University of California at Los Angeles, Los Angeles, Calif.); and **Arthur H. Hadley** (Robert M. Hadley Company, Inc., Los Angeles, Calif.): Since design is central to all engineering, it is hoped that this constructive paper will encourage similar efforts with respect to other probems, with subsequent publication of results. Our comments on the paper concern the computer application, and a few technical questions.

In connection with computer application, and the authors' introductory comments about the difference between small and large transformer design, the paper does not follow up sufficiently to demonstrate whether their computer program meets the standards they set up. The questions that come immediately to mind are of the following type: Given a customer specification, how long does it take to obtain a design? How much editing of the specification is required? What if the specification deviates slightly from the assumed conditions in the program? How do these over-all times compare with the over-all time to effect a design now?

In other words, it is observed in the conclusion that computing time is one second and operator time is one minute. No estimate is given for the preparation of the problem.

There is also the question of optimum design. The computer program appears to have been tailored with enough pretabulated data to effect a unique answer. Is would be of interest to learn what measures of "best" have been assumed and how well the computer design compares to these measures. Further, the question occurs of how does this design procedure differ from that practiced by humans? What has been lost or gained in the process of being forced to specify every step in the design?

In connection with design details, from the viewpoint of electronic transformers, the paper presents a logical approach to their design by means of digital computers. The sequence of steps (1 through 18) of the section "Simple 2-Winding Transformers" seems to be quite adequate.

In the section "Operating Frequencies" the statement is made "As indicated by Fig. 3, 400-cycle transformers weigh only a fraction of their 60-cycle counterparts, but have nearly the same regulation for a given rating." For equal va and identical ambients, and the final winding temperatures, a 400-cycle transformer will have considerably better regulation than a 60-cycle transformer. From the fact that, for equal va rating and equal temperature rise, a 400-cycle transformer is considerably smaller than a 60-cycle transformer, it can be seen that the over-all volume and surface area will be much less for a 400-cycle transformer. This means that in this smaller volume much less wattage can be generated. Less wattage, of course, means less copper loss and therefore better regulation. In the light of this condition the regulation curve of Fig. 3 should not give the same regulation at a given va for both 60-cycle and 400-cycle transformers if all other design criteria remain unchanged. It is, of course, probable that there has been a significant change in criteria which would account for similar 60- and 400-cycle regulation curves.

For example while it is stated earlier that the 60-cycle transformers were designed for 40 C rise and 65 C ambient, it is noted in the section "Notes on Programs" that the 400-cycle transformers are figured with an average ambient of 20 C and a maximum ambient of 70 C, while allowing a 80 C rise. This, coupled with Fig. 3, could cause some confusion as 400-cycle regulation is stated as being figured at 100 C. With further reference to Fig. 3, the weight in pounds should be stated as "core weight in pounds" as this checks with the core weight as shown in Fig. 2.

With a more complex computer programming, it can be seen that other types of electronic transformers such as audio transformers could be designed by digital computers. Of considerable future interest would be a computer program to examine the optimum values of the materials and data (i.e., core configurations, core materials, temperature rise, etc.) used as the foundations for the given design procedure.

L. F. Deise, W. Etchison, and **Reuben Lee:** Mr. Kirkwood's comments are appreciated. He manifests a good understanding of the value and use of computers for small electronic power transformers, which portends wider use of this design tool.

Several of the questions raised by Prof. Rosenstein and his associates are answered in the paper; the following will answer the remainder. Editing the customer's specifications has not been programed, nor does it appear practicable to foresee all the possible questions that may arise. At present, an engineer reviews all such specifications, regardless of whether the computer is used. With reasonable performance requirements, and even with deviations from assumed conditions, preparation of the problem is a matter of minutes. It is facilitated by the use of forms on which pertinent electrical values may be filled in. Depending upon the requirements and upon engineering efficiency, the total saving is anywhere from 1 to 2 days for a new transformer design.

For a given list of standard cores, the computer design depends wholly upon the assumed conditions. The design procedure follows closely that of a design engineer. Gains and losses, and the degree of optimization, are mentioned in the section "Simple 2-Winding Transformers." Weight minimization is limited only by the number of standard cores, and this in turn is an economic consideration. Weights shown in Fig. 3 are core weights.

Regulation in transformers for 400-cycle operation has been calculated at 100 C, compared with 75 C for 60 cycles. This, together with different cores, induction, current density, and maximum temperature rise, accounts for the percentage regulation's being nearly the same at both frequencies.

With due regard for the possibility of computer optimization in reducing weight from present practice, this is highly dependent on the properties of materials used. Computer designs described in this paper compare very favorably with other extant designs. Further progress hinges largely on improvements in transformer materials.

---

l. 79
1
60

The Synthesis of Linear Filters with Real or Imaginary Transfer Functions

ELMER G. GILBERT
NONMEMBER AIEE

JOSEPH OTTERMAN
NONMEMBER AIEE

THE SYNTHESIS of linear filters with symmetrical (even-function) or antisymmetrical (odd-function) impulse responses has both theoretical interest and practical value. Such filters which can be called "time-symmetric" or "time-antisymmetric," following the terminology of Storey and Grierson,[1] are characterized by transfer functions which are purely real or purely imaginary. Alternate names, proposed and employed in this paper, are real or imaginary filters. Applications of real or imaginary filters include compensating systems which must have no phase distortion, such as those often used in the field of acoustics and in the compensation of signals generated by picture scanning devices.[2,3]

In the past, little work of a general nature has been done. Although Kallmann,[4] Wiener and Lee,[5,6] and others[7,8] consider delay-line tranversal filters which have real or imaginary transfer functions, their emphasis is on transfer functions periodic in frequency. An excellent paper by Levy[9] describes some of the properties and the usefulness of real and imaginary transfer functions; however, it limits synthesis to a lumped-element delay-line filter with many experimentally adjusted reflections. Storey and Grierson[1] have used a tape-recorder scheme which involves storing the input signal on magnetic tape and processing the signal through reverse playback. Since these past contributions are restricted to limited types of transfer functions and physical realizations, their significance has not generally been appreciated.

It is the purpose of this paper to extend previous work by developing fully the conditions under which real and imaginary filters can be realized and by utilizing an orthogonal function approach for solving the approximation problem. Filter realizations presented in this paper do require a delay element. And although it is true that real or imaginary filters can be realized with arbitrarily small errors by means of conventional lumped-element filters, it is felt that the filters presented here can often be more simply and economically realized particularly when phase accuracy requirements are high.

Statement of the Problem

The problem of synthesis is to realize physically and as closely as practicable a prescribed mathematical operation an an input function of time. For a linear time invariant system the prescribed operation may be expressed by the superposition integral

$$f_0(t)=\int_0^{\infty} h(\tau)f_i(t-\tau)d\tau \tag{1}$$

where t is time, f_i is the input function, f_0 is the output function, and $h(t)$ is the response of the prescribed system to the unit impulse $u_0(t)$ occurring at $t=0$. The operation may also be expressed in terms of the Fourier transforms of f_i, f_0, and h by

$$F_0(j\omega)=H(j\omega)F_i(j\omega) \tag{2}$$

(The upper-case-letter functions denote the transforms of the lower-case-letter functions; i.e., $H(j\omega)=\int_{-\infty}^{\infty}\epsilon^{-j\omega t}h(t)\,dt$. It is tacitly assumed that the transforms exist when employed.) $H(j\omega)$ is the sinusoidal response function or the transfer function. The Fourier transform is used rather than the Laplace transform because impulse responses $h(t)$ will be considered which are nonzero for negative t. Synthesis is assumed complete when a physical system is determined which has an impulse response $h^*(t)$ which approximates $h(t)$ sufficiently well for a given purpose.

An approximation generally results because the physical system imposes the realizability condition, $h^*(t)=0$ for $t<0$, and possible practical conditions, such as being realized by a lumped-element system ($H(j\omega)$ rational in $j\omega$). In the present case either H_R or H_I (both real functions) in

$$H=H_R+jH_I \tag{3}$$

is zero. To approximate as closely as possible such an H, we will develop the conditions under which it is possible to realize an

$$H^*=H_R^*+jH_I^* \tag{4}$$

(H_R^* and H_I^* are real functions) where either $H_I^*=0$ or $H_R^*=0$, i.e., where the realized filter is either real or imaginary.

Conditions for Realizability

From the Fourier transform we have

$$H_R^*=\int_{-\infty}^{\infty} h^*(t)\cos\omega t\,dt$$

$$H_I^*=-\int_{-\infty}^{\infty} h^*(t)\sin\omega t\,dt \tag{5}$$

Thus it is clear that a symmetric impulse response

$$h^*(t)=h^*(-t) \tag{6}$$

Paper 60-202, recommended by the AIEE Basic Sciences Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., January 31–February 5, 1960. Manuscript submitted September 30, 1959; made available for printing January 8, 1960.

ELMER G. GILBERT is with the University of Michigan, Ann Arbor, Mich.; and JOSEPH OTTERMAN is with International Telephone and Telegraph Laboratories, Nutley, N. J.; located at Bell Telephone Laboratories, Inc., Whippany, N. J.

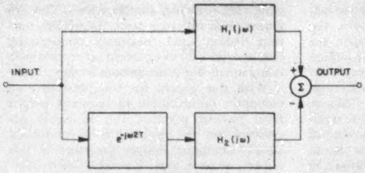


Fig. 1 (left). Single delay element realization of H*(jω)

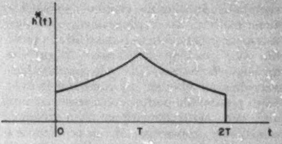


Fig. 2 (right). An h*(t) utilizing more than one delay element

yields a real filter ($H_I{}^*=0$) and an antisymmetric impulse response

$$h^*(t)=-h^*(-t) \tag{7}$$

yields an imaginary filter ($H_R{}^*=0$).

Since realizability requires $h^*(t)=0$ for $t<0$, it is clear from equations 6 and 7 that realizable real and imaginary filters require $h^*(t)=0$ except at $t=0$. The only functions which satisfy this condition are the unit impulse u_0 and its derivatives, u_1, u_2, ... But the impulse derivatives have no value as approximating functions and are not realizable by physical systems having finite gain at infinite frequency. Hence real and imaginary filters cannot be realized directly

By introducing a time delay T, certain real and imaginary filters can be realized. The time delay allows $h^*(t)$ to be symmetric or antisymmetric about $t=T$ instead of $t=0$. Realizability then means that

$$h^*(t)=0, \qquad t<0,\ t>2T \tag{8}$$

$$h^*(t+T)=h^*(-t+T), \qquad 0\le t\le T \tag{9}$$

(a delayed real filter)

$$h^*(t+T)=-h^*(-t+T) \qquad 0\le t\le T \tag{10}$$

(a delayed imaginary filter)

Fortunately, in most practical applications the added delay T does not limit the usefulness of these filters.

Since lumped-element systems have impulse responses of the form

$$h^*(t)=\sum_{n=1}^{N} c_n t^{k_n} e^{\alpha_n t}+c_0u_0(t),\ t\ge 0 \tag{11}$$

(k_n = zero or positive integer, α_n = real, imaginary, or complex number) and have infinite memory, it is obvious that they cannot satisfy equation 8 which states that $h^*(t)$ has a finite memory of $2T$. Only through the introduction of distributed elements may equations 8 and 9 or 8 and 10 be satisfied.

Physical Realization

Fig. 1 shows how a single delay-element transfer function $\epsilon^{-j\omega 2T}$ and two lumped-element transfer functions, $H_1(j\omega)$ and $H_2(j\omega)$, may be used to realize $H^*(j\omega)$. For $t\le 2T$ the impulse response is simply $h_1(t)$; for $t>2T$ the impulse response is $h_1(t)-h_2(t-2T)$. To make $h^*(t)=0$ for $t>2T$ it is necessary to make the responses of the two paths cancel for $t>2T$, i.e., to make $h_1(t)=h_2(t-2T)$ for $t>2T$. Thus

$$\begin{aligned} h^*(t)&=0, && t<0\\ &=h_1(t), && 0\le t\le 2T\\ &=0, && t>2T \end{aligned} \tag{12}$$

and

$$\begin{aligned} h_2(t)&=0, && t\le 0\\ &=h_1(t+2T), && t>0 \end{aligned} \tag{13}$$

Since h_1 must symmetric or antisymmetric about $t=T$ in $(0, 2T)$ it is clear that h_1 can have no terms with an exponential function of time as a factor. That is, the terms in h_1 must be of the following type:

Type of Functions	Form of Realization
$1, t, t^2, \ldots, t^k$	$(k+1)$ integrators
$\cos\omega_1 t$, $\sin\omega_1 t$, $\cos\omega_2 t$, $\sin\omega_2 t$, ...	undamped filter with poles at $j\omega=\pm j\omega_1$, $\pm j\omega_2$, ...
$\cos\omega_1 t$, $\sin\omega_1 t$, $t\cos\omega_1 t$, $t\sin\omega_1 t$, ..., $t^k\cos\omega_1 t$, $t^k\sin\omega_1 t$	undamped filter with $(k+1)$ order poles at $j\omega=\pm j\omega_1$

In every case the poles of H_1 (and also the poles of H_2 since by equation 13 the poles of H_2 must also be the poles of H_1) are on the imaginary axis in the complex p plane. In a later section approximating series based on the first two types of terms will be described.

Since passive circuit realization of H_1 or H_2 demands lossless inductors and capacitors, active circuitry is preferred for high accuracy realization. Actually, active circuit realization is mandatory where H_1 or H_2 are voltages ratios and have poles at $p=0$.

Other forms for $h^*(t)$ can be obtained if more delay elements are introduced. Fig. 2 shows an example. Here

$$h^*(t)=h_T(t)+h_{2T}(t) \tag{14}$$

where

$$\begin{aligned} h_T(t)&=0, && t<0, && t>T\\ h_{2T}(t)&=0, && t<T, && t>2T \end{aligned} \tag{15}$$

and

$$h_{2T}(t+T)=h_T(-t+T), \qquad 0\le t\le T \tag{16}$$

(a delayed real filter)

or

$$h_{2T}(t+T)=-h_T(-t+T), \qquad 0\le t\le T \tag{17}$$

(a delayed imaginary filter)

The functions h_T and h_{2T} are realized as $h^*(t)$ was realized in Fig. 1. In addition to requiring more than one delay element, the realization possesses practical limitations. If either h_T or h_{2T} contains a term with a decaying exponential factor, then by the symmetry conditions of equations 16 or 17 the other contains a term with a corresponding growing exponential factor which leads to an active unstable lumped-element filter in the over-all realization. Since the matching of gains in the unstable lumped-element filters cannot be exact, it is impossible to make $|h^*(t)|<\infty$ as $t\to\infty$. When the lumped element filters in h_T and h_{2T} have all their poles on the imaginary axis in the p plane a gain matching problem is still present, though it is not nearly so severe. Further remarks on accuracy of realization are given in the section titled "Illustrative Examples."

Many other types of responses are available from filters which contain both lumped elements and delay elements. Various authors have discussed limited classes of such filters.[10,11] For the realization of real and imaginary filters it does not seem worth-while to use more than a single delay element unless several delay elements will allow exact or nearly exact synthesis of the prescribed impulse response (at least within the time-delay T). Certainly, when $h(t)$ consists of sections of rather simple functions this is the case. Another example occurs when

$$h^*(t)=\sum_{n=1}^{N} a_n u_0(t-t_n) \tag{18}$$

t_n = real number

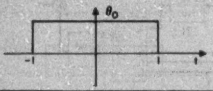
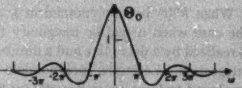
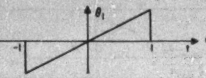
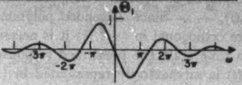
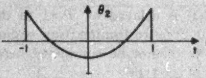
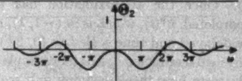
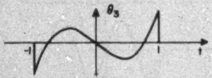
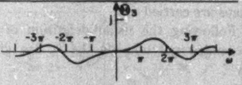
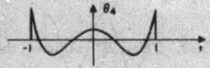
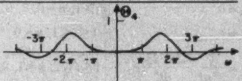
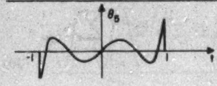
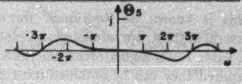
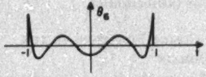
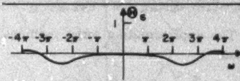
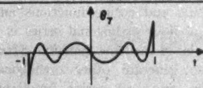
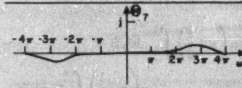
a_n = real number

Such an $h^*(t)$ can be realized with a tapped delay line and a summing amplifier and leads to the periodic transfer functions described by Kallmann,[4] Wiener and Lee,[5,6] White and Ruvin,[7] and Urkowitz.[8]

The Approximation Problem

Approximation involves the representation of a prescribed system function, $h(t)$

Table I

Time Domain		Frequency Domain	
	$\theta_0=1$	$\Theta_0=\frac{2\sin\omega}{\omega}$	
	$\theta_1=t$	$\Theta_1=j\left(-\frac{2\sin\omega}{\omega^2}+\frac{2\cos\omega}{\omega}\right)$	
	$\theta_2=\frac{1}{2}(3t^2-1)$	$\Theta_2=-\frac{6\sin\omega}{\omega^3}+\frac{6\cos\omega}{\omega^2}+\frac{2\sin\omega}{\omega}$	
	$\theta_3=\frac{1}{2}(5t_3-3t)$	$\Theta_3=j\left(\frac{30\sin\omega}{\omega^4}-\frac{30\cos\omega}{\omega^3}-\frac{12\sin\omega}{\omega^2}+\frac{2\cos\omega}{\omega}\right)$	
	$\theta_4=\frac{1}{8}(35t^4-30t^2+3)$	$\Theta_4=\frac{210\sin\omega}{\omega^5}-\frac{210\cos\omega}{\omega^4}-\frac{90\sin\omega}{\omega^3}+\frac{20\cos\omega}{\omega^2}+\frac{2\sin\omega}{\omega}$	
	$\theta_5=\frac{1}{8}(63t^5-70t^3+15t)$	$\Theta_5=j\left(\frac{-1890\sin\omega}{\omega^6}+\frac{1890\cos\omega}{\omega^5}+\frac{840\sin\omega}{\omega^4}-\frac{210\cos\omega}{\omega^3}-\frac{30\sin\omega}{\omega^2}+\frac{2\cos\omega}{\omega}\right)$	
	$\theta_6=\frac{1}{16}(231t^6-315t^4+105t^2-5)$	$\Theta_6=\frac{-20790\sin\omega}{\omega^7}+\frac{20790\cos\omega}{\omega^6}+\frac{9450\sin\omega}{\omega^5}-\frac{2520\cos\omega}{\omega^4}-\frac{420\sin\omega}{\omega^3}+\frac{42\cos\omega}{\omega^2}+\frac{2\sin\omega}{\omega}$	
	$\theta_7=\frac{1}{16}429t^7-693t^5+315t^3-35t)$	$\Theta_7=j\left(\frac{270270\sin\omega}{\omega^8}-\frac{270270\cos\omega}{\omega^7}-\frac{124740\sin\omega}{\omega^6}+\frac{34650\cos\omega}{\omega^5}+\frac{6300\sin\omega}{\omega^4}-\frac{756\cos\omega}{\omega^3}-\frac{56\sin\omega}{\omega^2}+\frac{2\cos\omega}{\omega}\right)$	

or $H(j\omega)$, in terms of a physically realizable system function ($h^*(t)$ or $H^*(j\omega)$). If the physical system is lumped, then $H^*(j\omega)$ is rational in $j\omega$; if the physical system has a delay real or imaginary transfer function, then $h^*(t)$ and $H^*(j\omega)$ assume the forms discussed in the previous section. Of the many approximating techniques available, this paper will present a minimum integral-square-error approach based on orthogonal approximating functions. Orthogonal function approximation has the following advantages: 1. The integral-square-error is minimized andeasily computed. 2. The approximation may be computed readily in either the time or frequency domains. 3. Approximation error tends to be small throughout the interval of expansion (as opposed to a Taylor expansion where the approximation tends to be good only at one point). 4. Orthogonal functions suitable for realization of real and imaginary filters are well known and tabulated. 5. Additional terms may be added to an approximating series to further reduce error without forcing a new calculation of the terms already determined. Orthogonal functions have been employed by many authors for the synthesis of lumped element systems.[12-16] However, because of the objectives of this paper, different types of approximating functions must be considered here. Whereas the orthogonal time functions which appear in the series for $h^*(t)$ for a lumped-element filter consist of exponential functions defined in the time interval $(0, \infty)$, the functions used in real or imaginary filter approximation must be orthogonal in $(0, 2T)$ and be symmetric or antisymmetric about $t=T$.

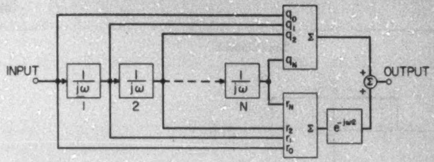


Fig. 3. Realization of real or imaginary filter based on Legendre function approximation

Orthogonal Approximating Functions

Legendre Polynomials

When $h^*(t)$ is a polynomial in t, as is the case when a real or imaginary filter is realized by a delay line and a number of integrators, Legendre polynomials are employed to form an orthogonal set $\theta_0(t)$, $\theta_1(t)$, ... Since Legendre polynomials are orthogonal in $(-1, 1)$ it is necessary to time scale the prescribed response so $h(t)$ is satisfactorily represented in $(-1, 1)$. Once the approximation has been completed, $h^*(t)$, which is in $(-1, 1)$, may be delayed and time scaled to $(0, 2T)$ for physical realization. A later example will illustrate how these delaying and scaling steps are carried out.

Following the usual definition of the Legendre polynomials the θ_n are given by[16]

$$\begin{aligned}\theta_n &= 0, && |t|>1,\ n=0, 1, 2, \ldots\\ &= 1, && -1\leq t\leq 1,\ n=0\\ &= \frac{1}{2^n n!}\,\frac{d^n(t^2-1)^n}{dt^n}, && -1\leq t\leq 1,\ n=1, 2, 3,\end{aligned} \tag{19}$$

This is known as Rodriques' formula. Several of the θ_n are tabulated and plotted in Table I. If higher order functions are required they can be obtained most conveniently from the recursion relation[16]

$$n\theta_n=(2n-1)t\theta_{n-1}-(n-1)\theta_{n-2} \tag{20}$$

The approximating series assumes the form

$$h^*(t)=\sum_{n=0}^{N} c_n\theta_n(t) \tag{21}$$

where

$$c_n=\left(n+\frac{1}{2}\right)\int_{-1}^{1} h\theta_n dt =\left(n+\frac{1}{2}\right)\int_{-\infty}^{\infty} h\theta_n dt \tag{22}$$

The factor $(n+{}^1/_2)$ occurs in equation 22' because the θ_n are not normalized, i.e., $\int_{-1}^{1}\theta_n{}^2dt=(n+{}^1/_2)^{-1}\neq 1$. Since N is finite (filter complexity grows directly with N) the integral-square error,

$$E=\int_{-\infty}^{\infty}(h-h^*)^2dt=\int_{-\infty}^{\infty}h^2dt-\sum_{n=0}^{N}\frac{c_n{}^2}{n+\frac{1}{2}} \tag{23}$$

is in general not zero. Approximation accuracy requirements set an upper bound on E which in turn establishes the minimum N.

From Table I it is seen that the θ_n are symmetric functions for even n and antisymmetric functions for odd n. Thus the even terms are used for a real filter approximation and the odd terms for an imaginary filter approximation. This can also be seen from equation 22 which gives $c_n=0$ for n odd when h is a symmetric function and $c_n=0$ for n even when h is an antisymmetric function.

In the frequency domain equations 21, 22, and 23 become respectively

$$H^*(j\omega)=\sum_{n=0}^{N} c_n\Theta_n(j\omega) \tag{24}$$

$$c_n=\frac{1}{2\pi}\left(n+\frac{1}{2}\right)\int_{-\infty}^{\infty} H\bar{\Theta}_n d\omega \tag{25}$$

and

$$E=\frac{1}{2\pi}\int_{-\infty}^{\infty}|H|^2d\omega-\sum_{n=0}^{N}\frac{c_n{}^2}{n+\frac{1}{2}} \tag{26}$$

where $\Theta_n(j\omega)$ is the Fourier transform of $\theta_n(t)$ and $\bar{\Theta}_n$ is the complex conjugate of Θ_n. Expressions 25 and 26 are derived from 22 and 23 by application of Parseval's theorem.[17]

The formulas for Θ_0 and Θ_1 are

$$\Theta_0=\int_{-\infty}^{\infty}\epsilon^{-j\omega t}\theta_0 dt=\int_{-1}^{1}\epsilon^{-j\omega t}dt=\frac{2\sin\omega}{\omega} \tag{27}$$

and

$$\Theta_1=\int_{-\infty}^{\infty}\epsilon^{-j\omega t}\theta_1 dt=\int_{-1}^{1}\epsilon^{-j\omega t}t\,dt =j2\left[\frac{\cos\omega}{\omega}-\frac{\sin\omega}{\omega^2}\right] \tag{28}$$

Although the higher order functions can be obtained in the same way, a much more rapid procedure is to use the recursion formula

$$\Theta_n=\frac{2n-1}{j\omega}\Theta_{n-1}+\Theta_{n-2} \tag{29}$$

which is developed in the Appendix. The Θ_n are also tabulated and plotted in Table I. As should be expected, the Θ_n are real for even n and imaginary for odd n.

Table I shows that the approximation H^* contains terms of the type

$$\frac{j\sin\omega}{(j\omega)^n}=\frac{\epsilon^{j\omega}-\epsilon^{-j\omega}}{2(j\omega)^n}$$

and

$$\frac{\cos\omega}{(j\omega)^n}=\frac{\epsilon^{j\omega}+\epsilon^{-j\omega}}{2(j\omega)^n}$$

When $h^*(t)$ is delayed by $T=1$ to the interval $(0, 2)$ for realization (time scaling to $(0, 2T)$ will not be considered here) functions of the type

$$\frac{1-\epsilon^{-j\omega 2}}{2(j\omega)^n}$$

and

$$\frac{1+\epsilon^{-j\omega 2}}{2(j\omega)^n}$$

are obtained. Thus H^* assumes the form

$$H^*=\sum_{n=0}^{N} q_n\frac{1}{(j\omega)^n}+\epsilon^{-j\omega 2}\sum_{n=0}^{N} r_n\frac{1}{(j\omega)^n} \tag{30}$$

where the q_n and r_n are determined from the c_n and the expressions for the Θ_n. Fig. 3 shows the realization.

Sine and Cosine Functions in the Time Domain

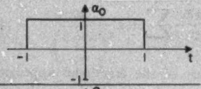
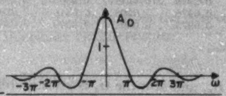
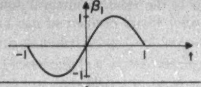
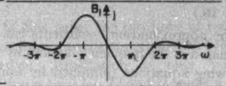
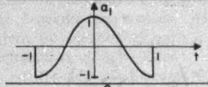
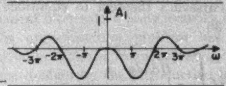
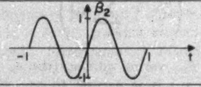
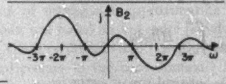
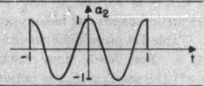
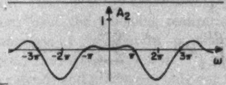
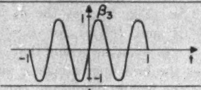
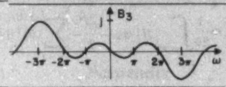
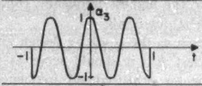
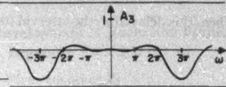
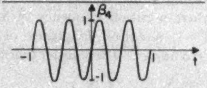
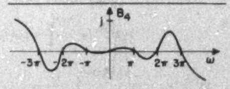
If a real or imaginary filter is realized with a single delay line and an undamped lumped element filter, h^* consists of sections of sine and cosine functions and the most obvious orthogonal series is a Fourier series in the time domain. (It is possible to generate other orthogonal series from sine and cosine functions. For example the sine and cosine functions would not have to be harmonic.) For convenience we will consider a Fourier series with the interval of expansion $(-1, 1)$. Then the previous remarks which were made concerning time scale still hold.

Since the period of expansion is 2 the Fourier series can be written as

$$\begin{aligned}h^*&=\frac{a_0}{2}+\sum_{n=1}^{N}(a_n\cos\pi nt+b_n\sin\pi nt), && -1\leq t\leq 1\\ &=0, && |t|>1\end{aligned} \tag{31}$$

Because two types of approximating functions are involved for each n ($\sin\pi nt$ and $\cos\pi nt$) the functions

Table II

Time Domain		Frequency Domain	
	$\alpha_0(t)=1$	$A_0(\omega)=\frac{2\sin\omega}{\omega}$	
	$\beta_1(t)=\sin\pi t$	$B_1(\omega)=-j\frac{2\pi\sin\omega}{\pi^2-\omega^2}$	
	$\alpha_1(t)=\cos\pi t$	$A_1(\omega)=+\frac{2\omega\sin\omega}{\pi^2-\omega^2}$	
	$\beta_2(t)=\sin 2\pi t$	$B_2(\omega)=+j\frac{4\pi\sin\omega}{4\pi^2-\omega^2}$	
	$\alpha_2(t)=\cos 2\pi t$	$A_2(\omega)=-\frac{2\omega\sin\omega}{4\pi^2-\omega^2}$	
	$\beta_3(t)=\sin 3\omega t$	$B_3(\omega)=-j\frac{6\pi\sin\omega}{9\pi^2-\omega^2}$	
	$\alpha_3(t)=\cos 3\pi t$	$A_3(\omega)=+\frac{2\omega\sin\omega}{9\pi^2-\omega^2}$	
	$\beta_4(t)=\sin 4\pi t$	$B_4(\omega)=+j\frac{8\pi\sin\omega}{16\pi^2-\omega^2}$	

$$\begin{aligned}\alpha_n(t)&=\cos\pi nt, && -1\le t\le 1\\ &=0, && |t|>1\end{aligned}\qquad(32)$$

$$\begin{aligned}\beta_n(t)&=\sin\pi nt, && -1\le t\le 1\\ &=0, && |t|>1\end{aligned}\qquad(33)$$

are defined. The Fourier series in equation 31 then becomes

$$h^*(t)=\frac{a_0}{2}\alpha_0+\sum_{n=1}^{N}a_n\alpha_n+\sum_{n=1}^{N}b_n\beta_n\qquad(34)$$

where the coefficients are given by

$$a_n=\int_{-\infty}^{\infty}h\alpha_n dt=\int_{-1}^{1}h\cos\pi nt\,dt\qquad(35)$$

and

$$b_n=\int_{-\infty}^{\infty}h\beta_n dt=\int_{-1}^{1}h\sin\pi nt\,dt\qquad(36)$$

For a real filter $b_n=0$ and for an imaginary filter $a_n=0$. Thus the α_n form real filter approximations and the β_n form imaginary filter approximations.

In the frequency domain the real-filter functions are given by the Fourier transform of the α_n

$$\begin{aligned}A_n&=\int_{-\infty}^{\infty}\epsilon^{-j\omega t}\alpha_n dt=\int_{-1}^{1}\epsilon^{j\omega t}\cos\pi nt\,dt\\ &=\frac{j\omega(-1)^n(\epsilon^{j\omega}-\epsilon^{-j\omega})}{(j\omega)^2+(\pi n)^2}\\ &=-\frac{2\omega(-1)^n\sin\omega}{(j\omega)^2+(\pi n)^2}\end{aligned}\qquad(37)$$

Similarly for the imaginary-filter functions

$$\begin{aligned}B_n&=\int_{-\infty}^{\infty}\epsilon^{-j\omega t}\beta_n dt=\int_{-1}^{1}\epsilon^{-j\omega t}\sin\pi nt\,dt\\ &=\frac{\pi n(-1)^n(\epsilon^{j\omega}-\epsilon^{-j\omega})}{(j\omega)^2+(\pi n)^2}\\ &=j\frac{2\pi n(-1)^n\sin\omega}{(j\omega)^2+(\pi n)^2}\end{aligned}\qquad(38)$$

Table II summarizes these results and shows plots of the functions α_n, β_n, A_n, and B_n. (A brief tabulation of the A_n

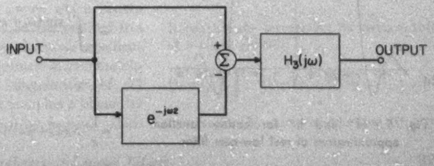


Fig. 4. Realization of real or imaginary filter based on Fourier series approximation

Table III. E_{rel} for Cosine Function Approximation of Real Low Pass Filter, ω_c

N	π	2π	3π	3.5π	4π	4.5π	5π	6π	7π
3	0.0978	0.0511	0.0363	0.0307	0.0869	0.2176	0.3125		
4	0.0976	0.0506	0.0345	0.0287	0.0275	0.0230	0.0702	0.2600	
5	0.0975	0.0504	0.0340	0.0286	0.0260	0.0221	0.0222	0.0589	0.2226

and B_n, and also the Θ_n is given in reference 18.)

The approximation coefficients can be evaluated in the frequency domain by the following equations (obtained by applying Parseval's theorem to equations 35 and 36

$$a_n=\frac{1}{2\pi}\int_{-\infty}^{\infty} H\bar{A}_n d\omega \tag{39}$$

$$b_n=\frac{1}{2\pi}\int_{-\infty}^{\infty} H\bar{B}_n d\omega \tag{40}$$

The integral-square-error is

$$\begin{aligned}E&=\int_{-1}^{1} h^2dt-\frac{a_0^2}{2}-\sum_{n=1}^{N}(a_n^2+b_n^2)\\&=\frac{1}{2\pi}\int_{-\infty}^{\infty}|H|^2d\omega-\frac{a_0^2}{2}-\sum_{n=1}^{N}(a_n^2+b_n^2)\end{aligned} \tag{41}$$

When h^* is delayed to the interval (0, 2) for realization the functions A_n and B_n become

$$A_n\epsilon^{-j\omega}=\frac{j\omega(-1)^n(1-\epsilon^{-j\omega 2})}{(j\omega)^2+(\pi n)^2} \tag{42}$$

and

$$B_n\epsilon^{-j\omega}=\frac{\pi n(-1)^n(1-\epsilon^{-j\omega 2})}{(j\omega)^2+(\pi n)^2} \tag{43}$$

Thus H^* assumes the form

$$H^*=(1-\epsilon^{-j\omega 2})H_3 \tag{44}$$

and may be realized as shown in Fig. 4. For a real filter

$$H_2=\frac{a_0}{2j\omega}+\sum_{n=1}^{N}\frac{a_n(j\omega)(-1)^n}{(j\omega)^2+(\pi n)^2} \tag{45}$$

For an imaginary filter

$$H_3=\sum_{n=1}^{N}\frac{b_n\pi n(-1)^n}{(j\omega)^2+(\pi n)^2} \tag{46}$$

H_3 can be realized as a voltage ratio with lossless passive circuits except for the $a_0/2j\omega$ term.

Illustrative Examples

For the first example a real, ideal low-pass filter is approximated in the time interval $(-1, 1)$ by sine and cosine functions. If the ideal filter has a cut-off frequency ω_c, then the transfer function is

$$\begin{aligned}H=H_R&=1, &&-\omega_c\le\omega\le\omega_c\\&=0, &&|\omega|>\omega_c\end{aligned} \tag{47}$$

and

$$h=\frac{\sin\omega_c t}{\pi t} \tag{48}$$

The coefficients a_n (the $b_n=0$) are determined most easily from equation 35 with h given by equation 48

$$\begin{aligned}a_n&=\int_{-1}^{1}\frac{\sin\omega_c t}{\pi t}\cos\pi nt\,dt=2\int_0^1\frac{\sin\omega_c t}{\pi t}\times\cos\pi nt\,dt\\&=\frac{1}{\pi}\int_0^1\frac{\sin(\omega_c+\pi n)t}{t}dt+\frac{1}{\pi}\int_0^1\frac{\sin(\omega_c-\pi n)t}{t}dt\\&=\frac{1}{\pi}[S_i(\omega_c+\pi n)+S_i(\omega_c-\pi n)]\end{aligned} \tag{49}$$

$S_i(x)=\int_0^x \sin y/y\,dy$, the sine integral function, for which tables are available. The integral-square error is given by equation 41 as

$$E=\frac{\omega_c}{\pi}-\frac{a_0^2}{2}-\sum_{n=1}^{N}a_n^2 \tag{50}$$

A measure of the fractional approximation error is the relative integral square error

$$E_{rel}=\frac{E}{\int_{-\infty}^{\infty}h^2dt} \tag{51}$$

which in this case is

$$E_{rel}=1-\frac{\pi}{\omega_c}\left(\frac{a_0^2}{2}+\sum_{n=1}^{N}a_n^2\right) \tag{52}$$

Since E_{rel} varies with ω_c (the a_n depend on ω_c) it is natural to ask what cut-off frequency yields the best approximation. Table III gives some results for $N=3$, 4, 5. For $N=3$ (four nonzero terms in h^*) the smallest E_{rel} is obtained for $\omega_c=3.5\pi$. Fig. 5 gives the a_n and shows $h^*(t)$ and $H^*(j\omega)$ for $\omega_c=3.5\pi$ and $N=3$. It is not surprising that $\omega_c=3.5\pi$ gives the best approximation since $3.5\pi\cong 3\pi$, the highest frequency component in $h^*(t)$ for $N=3$.

Inasmuch as the $\omega_c=3.5\pi$ design gives the smallest E_{rel} it should serve as the prototype for other cut-off frequencies. Thus, where a cut-off frequency of 3.5π yields $T=1$, a cut-off frequency of ω_c yields $T=3.5\pi/\omega_c$. The transfer function is

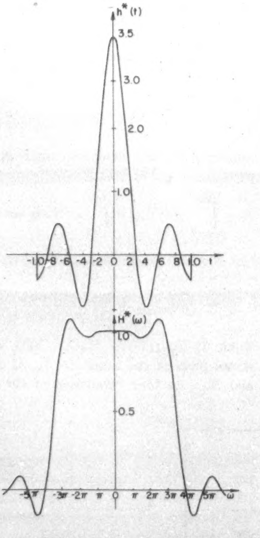


Fig. 5. H* and h* for cosine function approximation of real low-pass filter

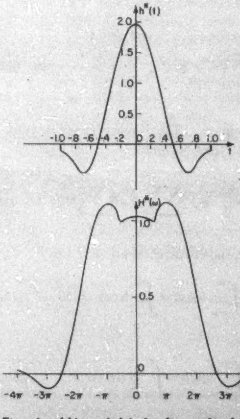


Fig. 6. H* and h* for Legendre function approximation of real low-pass filter

$$H^*=2\left(\frac{3.5\pi\omega}{\omega_c}\right)\sin\frac{3.5\pi\omega}{w_c}\times$$
$$\left[0.5025\left(\frac{\omega_c}{3.5\pi\omega}\right)^2+\frac{-0.9937}{\pi^2-\left(\frac{3.5\pi\omega}{\omega_c}\right)^2}+\right.$$
$$\left.\frac{1.0130}{4\pi^2-\left(\frac{3.5\pi\omega}{\omega_c}\right)^2}+\frac{-0.9356}{9\pi^2-\left(\frac{3.5\pi\omega}{\omega_c}\right)^2}\right] \tag{53}$$

and can be realized in the form of Fig. 4 with an added delay $T=3.5\pi/\omega_c$.

Proceeding in a similar fashion the real, ideal low pass filter can be approximated by Legendre polynomials. The c_n are given by equation 22 with h given by equation 48 and θ_n by Table I. Of course, $c_n=0$ for odd n. Equation 22 involves integrals of the form $\int_0^1 t^k \sin\omega_c t\ dt$ ($k=-1, 1, 3, 5\ldots$) which are readily evaluated. From equations 23 and 51

$$E_{rel}=1-\frac{\pi}{\omega_c}\sum_{n=0}^{N}\frac{1}{n+\frac{1}{2}}c_n^2 \tag{54}$$

In this case the smallest E_{rel} is obtained for $\omega_c=2\pi$ when $N=6$ (four nonzero terms in h^*). Fig. 6 gives the c_n and shows $h^*(t)$ and $H^*(j\omega)$ for $\omega_c=2\pi$ and $N=6$. Note that $E_{rel}=0.0506$ which is somewhat higher than obtained for the same number of terms in the cosine function representation. Using the forgoing design for an $N=6$ prototype yields $T=2\pi/\omega_c$ and the transfer function

$$H^*=\sin\frac{2\pi\omega}{\omega_c}\left[-0.1988\frac{\omega_c}{2\pi\omega}+\right.$$
$$115.6\left(\frac{\omega_c}{2\pi\omega}\right)^3-5{,}195\left(\frac{\omega_c}{2\pi\omega}\right)^5+$$
$$\left.12{,}140\left(\frac{\omega_c}{2\pi\omega}\right)^7\right]+\cos\frac{2\pi\omega}{\omega_c}\times$$
$$\left[-2.747\left(\frac{\omega_c}{2\pi\omega}\right)^2+1{,}148\left(\frac{\omega_c}{2\pi\omega}\right)^4-\right.$$
$$\left.12{,}140\left(\frac{\omega_c}{2\pi\omega}\right)^6\right] \tag{55}$$

Realization in the form of Fig. 3 requires an additional delay $T=2\pi/\omega_c$

The last example is a sine and cosine function approximation of the imaginary, ideal low-pass filter characteristic

$$\begin{aligned}H=jH_I&=+j, & 0<\omega\le\omega_c\\ &=-j, & -\omega_c\le\omega<0\\ &=0, & \omega=0,\ |\omega|>\omega_c\end{aligned} \tag{56}$$

and

$$h=\frac{\cos\omega_c t-1}{\pi t} \tag{57}$$

The coefficients b_n (the $a_n=0$) are given from equation 36 and 57 by

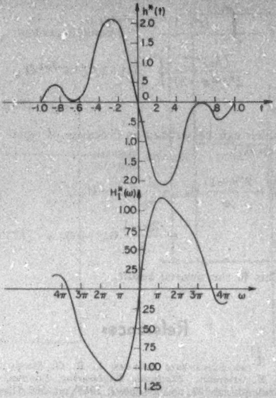


Fig. 7. H* and h* for sine function approximation of imaginary low-pass filter

$$b_n=2\int_0^1\frac{\cos\omega_c t-1}{\pi t}\sin\pi n t\ dt$$
$$=\frac{1}{\pi}[S_i(\pi n+\omega_c)-2S_i(\pi n)+S_i(\pi n-\omega_c)] \tag{58}$$

and the relative mean square error by

$$E_{rel}=1-\frac{\pi}{\omega_c}\sum_{n=1}^{N}b_n^2 \tag{59}$$

In this case, E_{rel} is near its smallest value at $\omega_c=3\pi$ when $N=4$ (four nonzero terms in h^*) where $E_{rel}=0.119$. The relatively large E_{rel} is probably caused by the discontinuity in H at $\omega=0$ in addition those at $\omega=\pm\omega_c$. Fig. 7 gives the b_n and shows $h^*(t)$ and $HI^*(j\omega)$ for $\omega_c=3\pi$ and $N=4$.

Both of the realizations in Figs. 3 and 4 have parallel signal paths, one of which is direct and a second which is through a delay line. Gain matching in the two paths must be exact or cancellation of $h_1(t)$ and $h_2(t-2T)$ will not occur for $t>2T$. In the case of the Legendre realization any difference between $h_1(t)$ and $h_2(t-2T)$ grows as t^N which cause a large error in the realized frequency response at zero frequency; in the case of sines or cosines realization the difference between $h_1(t)$ and $h_2(t-2T)$ oscillates with small amplitude at frequencies $\omega=\pi n$ which causes undesired resonances, in the realized frequency response. A practical expedient for reducing the criticalness in gain matching is to shift the singularities of H^* slightly to the left in complex p plane (by replacing $j\omega$ in $H^*(j\omega)$ by $j\omega+\sigma_0$). This causes $h^*(t)$ to become multiplied by $\epsilon^{-\sigma_0 t}$, a damping factor which is effective in reducing the error for large t. The shift σ_0 should be made as small as possible consistent with the obtainable accuracy in gain matching as the shift causes H^* to depart from the desired real or imaginary characteristics.

Conclusions

In this paper the synthesis of real and imaginary transferfunctions has been described. Realization is only possible within a time lag T, and requires a delay element filter with a finite memory of $2T$. The approximation problem has been solved in both the frequency domain and time domain by the introduction of two sets of orthogonal approximating functions, one based on Legendre polynomials in the time domain and the other on sines and cosines in the time domain. It appears that the latter approximating functions have the following realization advantages: 1. realization is simple in form, consisting of $(1-\epsilon^{-j\omega 2T})$ and H_3 in cascade and 2. at most, only one pole is required at $p=0$.

Appendix

Development of Recursion Formula for Legendre Function $\Theta_n(j\omega)$

From equation 19

$$\Theta_n(j\omega)=\int_{-\infty}^{\infty}\epsilon^{-j\omega t}\Theta_n dt$$
$$=\frac{1}{2^n n!}\int_{-1}^{1}\frac{d^n(t^2-1)^n}{dt^n}\epsilon^{-j\omega t}dt \tag{60}$$

Integration by parts of equation 60 gives (making the derivative of $\epsilon^{-j\omega t}$ appear in the new integrand)

$$\Theta_n(j\omega)=\frac{1}{2^n n!}\left\{\frac{d^{n-1}(t^2-1)^n}{dt^{n-1}}\epsilon^{-j\omega t}\right\}\Bigg|_{-1}^{1}+$$
$$j\omega\int_{-1}^{1}\frac{d^{n-1}(t^2-1)^n}{dt^{n-1}}\epsilon^{-j\omega t}dt \tag{61}$$

but

$$\frac{d^{n-1}(t^2-1)^n}{dt^{n-1}}=0\text{ for }t=\pm 1$$

Hence

$$\Theta_n=j\omega\int_{-1}^{1}\frac{d^{n-1}(t^2-1)^n}{dt^{n-1}}\epsilon^{-j\omega t}dt \tag{62}$$

Repeating the integration by parts a total of n times yields

$$\Theta_n=\frac{(j\omega)^n}{2^n n!}\int_{-1}^{1}(t^2-1)^n\epsilon^{-j\omega t}dt \tag{63}$$

Integration by parts of equation 63 gives (making the integral of $\epsilon^{-j\omega t}$ appear in the new integrand)

$$\Theta_n=\frac{(j\omega)^n}{2^n n!}\left\{(t^2-1)^n\frac{\epsilon^{-j\omega t}}{(-j\omega)}\right\}\Bigg|_{-1}^{1}+$$

$$\frac{(j\omega)^{n-1}}{2^{n-1}(n-1)!}\int_{-1}^{1}(t^2-1)^{n-1}\,t\epsilon^{-j\omega t}\,dt$$

$$=\frac{(j\omega)^{n-1}}{2^{n-1}(n-1)!}\int_{-1}^{1}(t^2-1)^{n-1}t\epsilon^{-j\omega t}\,dt,\qquad n>0 \quad (64)$$

Finally, a similar integration by parts of equation 64 gives

$$\Theta_n=\frac{(j\omega)^{n-1}}{2^{n-1}(n-1)!}\left\{(t^2-1)^{n-1}\,t\,\frac{\epsilon^{-j\omega t}}{(-j\omega)}\right\}\Bigg|_{-1}^{1}+$$

$$\frac{2(n-1)}{j\omega}\int_{-1}^{1}(t^2-1)^{n-2}\,t^2\epsilon^{-j\omega t}\,dt+$$

$$\frac{1}{j\omega}\int_{-1}^{1}(t^2-1)^{n-1}\,\epsilon^{-j\omega t}\,dt$$

$$=\frac{(j\omega)^{n-2}}{2^{n-2}(n-2)!}\int_{-1}^{1}(t^2-1)^{n-2}\,t^2\epsilon^{-j\omega t}\,dt+$$

$$\frac{(j\omega)^{n-2}}{2^{n-1}(n-1)!}\int_{-1}^{1}(t^2-1)^{n-1}\,\epsilon^{-j\omega t}\,dt,\qquad n>1 \quad (65)$$

Adding and subtracting

$$\frac{(j\omega)^{n-2}}{2^{n-2}(n-2)!}\int_{-1}^{1}(t^2-1)^{n-2}(t^2-1)\epsilon^{-j\omega t}\,dt,$$

$$\Theta_n=\frac{(j\omega)^{n-2}}{2^{n-2}(n-2)!}\int_{-1}^{1}(t^2-1)^{n-2}(t^2-1)\times$$

$$\epsilon^{-j\omega t}\,dt+\frac{j\omega^{n-2}}{2^{n-1}(n-1)!}\times$$

$$\int_{-1}^{1}(t^2-1)^{n-1}\,\epsilon^{-j\omega t}dt+$$

$$\frac{(j\omega)^{n-2}}{2^{n-2}(n-2)!}\int_{-1}^{1}(t^2-1)^{n-2}\,\epsilon^{-j\omega t}\,dt \quad (66)$$

which can be written as (because of equation 63)

$$\Theta_n=\frac{2(n-1)}{j\omega}\,\Theta_{n-1}+\frac{1}{j\omega}\,\Theta_{n-1}+\Theta_{n-2}$$

$$=\frac{2n-1}{j\omega}\,\Theta_{n-1}+\Theta_{n-2} \quad (67)$$

This is the desired result.

References

1. TIME SYMMETRIC FILTERS, **L. R. O. Storey, J. K. Grierson.** *Electronic Engineering*, London, England, vol. 30, no. 568, pt. I, 1958, pp. 586–92; vol. 30, no. 569, pt. II, 1958, pp. 648–53.

2. APERTURE COMPENSATION FOR TELEVISION CAMERAS, **R. C. Dennison.** *RCA Review*, Princeton, N. J., Dec. 1953, pp. 569–85.

3. APERTURE CORRECTIVE SYSTEMS, **Joseph Otterman.** *Transactions*, Professional Group on Instrumentation, Institute of Radio Engineers, New York, N. Y., vol. PGI-8, no. 1, Mar. 1959, pp. 8–19.

4. TRANSVERSAL FILTERS, **Heinz E. Kallmann.** *Proceedings*, Institute of Radio Engineers, vol. 28, July 1940, pp. 302–08.

5. ELECTRICAL NETWORKS SYSTEM, **Y. W. Lee, N. Wiener.** *U. S. Patent no. 2,128,257*, filed July 7, 1936.

6. ELECTRICAL NETWORKS SYSTEM, **N. Wiener, Y. W. Lee.** *U. S. Patent no. 2,124,599*, filed July 18, 1936.

7. RECENT ADVANCES IN THE SYNTHESIS OF COMB FILTERS, **W. D. White, A. E. Ruvin.** *Convention Record*, Institute of Radio Engineers, vol. 5, pt. 2, 1957, pp. 186–99.

8. ANALYSIS AND SYNTHESIS OF DELAY LINE PERIODIC FILTERS, **Harry Urkowitz.** *Transactions*, Professional Group on Circuit Theory, Institute of Radio Engineers, vol. PGCT-4, June, 1957, pp. 41–53.

9. THE IMPULSE RESPONSE OF ELECTRICAL NETWORKS, WITH SPECIAL REFERENCE TO THE USE OF ARTIFICIAL LINES IN NETWORK DESIGN, **M. Levy.** *Journal*, Institution of Electrical Engineers, London, England, vol. 90, pt. III, no. 12, Dec., 1943, pp. 153–64.

10. FILTERS FOR SAMPLED SIGNALS, **J. Brogan.** *Proceedings*, Symposium on Information Networks, Polytechnic Institute of Brooklyn, Brooklyn, N.Y., Apr. 1954, pp. 71–83.

11. THE CONTINUOUS DELAY-LINE SYNTHESIZER AS A SYSTEM ANALOGUE, **J. H. Westcott.** *Monograph no. 176M*, Institution of Electrical Engineers, May 1956.

12. SYNTHESIS OF ELECTRICAL NETWORKS BY MEANS OF FOURIER TRANSFORMS OF LAGUERRE'S FUNCTIONS, **Y. W. Lee.** *Journal of Mathematics and Physics*, Cambridge, Mass., vol. II, 1932, pp. 83–113.

13. TRANSIENT SYNTHESIS IN THE TIME DOMAIN, **W. H. Kautz.** *Transactions*, Professional Group on Circuit Theory, Institute of Radio Engineers, vol. PGCT-1, Sept. 1954, pp. 29–39.

14. TIME DOMAIN NETWORK SYNTHESIS FOR AN ANALOG COMPUTER SETUP, **Joseph Otterman.** *Proceedings*, National Simulation Conference, Dallas, Tex., Jan. 1956, pp. 24.1–.5.

15. LINEAR SYSTEM APPROXIMATION BY DIFFERENTIAL ANALYZER SIMULATION OF ORTHONORMAL APPROXIMATING FUNCTIONS, **Elmer G. Gilbert.** *Transactions*, Professional Group on Electronic Computers, Institute of Radio Engineers, vol. PGEC-8, no. 2, June, 1959, pp. 204–09.

16. METHODS OF MATHEMATICAL PHYSICS (book), **R. Courant, D. Hilbert.** Interscience Publishers, New York, N. Y., vol. 1, 1953.

17. INTRODUCTION TO THE THEORY OF FOURIER INTEGRALS (book), **E. E. Titchmarsh.** Oxford University Press, London, England, 1937.

18. *Report S-58*, Radiation Laboratory, Massachusetts Institute of Technology, Cambridge, Mass., Aug. 1945.

Vol.
Pt. 1
196

Design of an Eddy-Current Brake for a Sodium-Cooled Nuclear Power Reactor

R. S. BAKER
MEMBER AIEE

THE sodium-cooled nuclear power reactor, known as the sodium reactor experiment (SRE), is operated by Atomics International, a division of North American Aviation, Inc., for the Atomic Energy Commission. The reactor is located in the Santa Susana Mountains near Los Angeles, Calif. The design features of this reactor have been outlined by Siegel.[1]

During the early operation of this reactor, it was determined that some form of sodium-flow throttling device was necessary to prevent excessive thermal stresses from occurring in the reactor core and sodium piping after a shutdown, or "scram."

When the reactor is shut down, the production of heat within it reduces to a small fraction of its former value. This small amount of heat (due to afterglow) that is still being produced is sufficient to cause circulation of sodium into and out of the reactor by thermal convection. The result is that sodium, at a relatively low temperature, is brought into contact with piping which is still at a high temperature, thereby leading to excessive thermal stresses. These stresses can be minimized by controlling the sodium flow in proportion to the heat generation, resulting in constant temperature gradients.

Method Used to Control Flow after Shutdown

For the purpose of throttling the sodium flow throughout the reactor piping to maintain a constant reactor temperature gradient for afterglow power conditions, two eddy-current brakes were installed in the SRE. One is used on the primary piping system, the other on the secondary system. Each is constructed as shown in Fig. 1.

Both brakes are controlled by thermo-

Paper 60-150, recommended by the AIEE Nucleonics Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., January 31–February 5, 1960. Manuscript submitted June 27, 1958; made available for printing December 8, 1959.

R. S. BAKER is with Atomics International, Division of North American Aviation, Inc., Canoga Park, Calif.

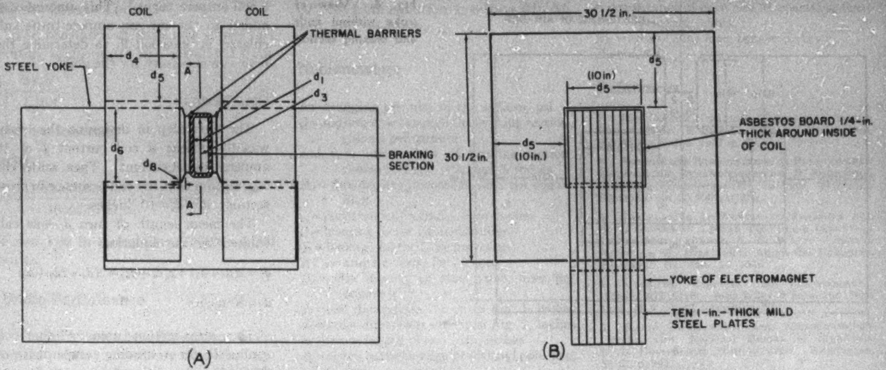

COIL
COIL
THERMAL BARRIERS
STEEL YOKE
d_4
d_5
A
d_1
d_3
d_6
d_8
BRAKING SECTION
A
(A)
30 1/2 in.
d_5
(10 in)
d_5
30 1/2 in.
d_5
(10 in.)
ASBESTOS BOARD 1/4-in. THICK AROUND INSIDE OF COIL
YOKE OF ELECTROMAGNET
TEN 1-in.-THICK MILD STEEL PLATES
(B)

Fig 1. Eddy-current brake. A—Cross section of braking section. B—Side view

couples that measure sodium temperatures in primary and secondary systems. Performance has been satisfactory. The braking pressures obtained at different sodium flow rates agree with calculated figures.

Design Calculations

DETERMINATION OF REQUIRED FLUX DENSITY

To begin the design of the d-c electromagnet which constitutes the brake, a basic configuration was assumed; see Fig. 1. Dimensions of the braking section and thermal barriers are shown in Fig. 2, where the symbols represent quantities that will be used later in the design calculations. Fig. 3 shows dimensions of the braking section and pole face in planes parallel to the flow direction. Fig. 4 shows the yoke without coils and braking section.

The calculation of magnet design to produce a given braking pressure is based on Rudenberg's equation for the eddy-current drag on a copper disk.[2,3] Written in terms of the quantities used in designing the eddy-current brakes, this equation becomes

$$P=\frac{8.5\times10^{-16}\,QB^2_G}{r_1d_1}\times\left[\frac{\dfrac{d_3}{d_4}+\dfrac{d_4}{d_3}}{\left(\dfrac{39.2\times10^{-9}Q}{r_1d_2}\right)^2+\left(\dfrac{d_3}{d_4}+\dfrac{d_4}{d_3}\right)^2}\right] \qquad (1)$$

In test runs conducted at the SRE, it was determined that under scram conditions the sodium flow rate in the main primary circuit, through the reactor, should drop to about 12 gallons per minute. It was also determined that the eddy-current brake should cause a dragging pressure of 0.3 psi (pounds per square inch) at this flow rate. The flux density necessary to produce this drag pressure was calculated from equation 1, for a sodium temperature of 1,000 degrees Fahrenheit.

$P=0.3$ psi
$d_1=2$ inches
$d_2=4$ inches
$d_3=7^3/_4$ inches
$d_4=10$ inches (this value was determined by adding an estimated amount of 2.25 inches for pole face chamber to d_3.)
$Q=12$ gallons per minute
$r_1=10.5\times10^{-6}$ ohm-inches

Solving for B_G yields $B_G=35{,}800$ lines per square inch.

CALCULATION OF MAGNETOMOTIVE FORCE REQUIRED

The next step in the design was the calculation of the coil ampere-turns required to set up this flux density. Two coils were used, one on each side of the air gap, and placed as close as possible to the gap. Each coil was proportioned to have as nearly as possible the maximum ratio of inductance to resistance. This requires that each coil be square in cross section, with a mean diameter which is 3.02 times the length of one side;[4,5] see Fig. 1. Because of space limitations, it was not possible to construct the magnet yoke to meet this requirement precisely; however, the actual performance has not been measurably affected.

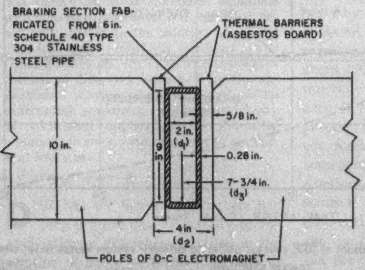

BRAKING SECTION FABRICATED FROM 6 in. SCHEDULE 40 TYPE 304 STAINLESS STEEL PIPE
THERMAL BARRIERS (ASBESTOS BOARD)
10 in.
9 in
2 in. (d_1)
5/8 in.
0.28 in.
7-3/4 in. (d_3)
4 in (d_2)
POLES OF D-C ELECTROMAGNET

Fig. 2. Cross section of braking section

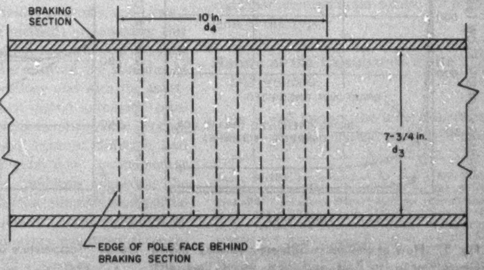

BRAKING SECTION
10 in. d_4
7-3/4 in. d_3
EDGE OF POLE FACE BEHIND BRAKING SECTION

Fig 3. Section A-A of Fig. 1(A)

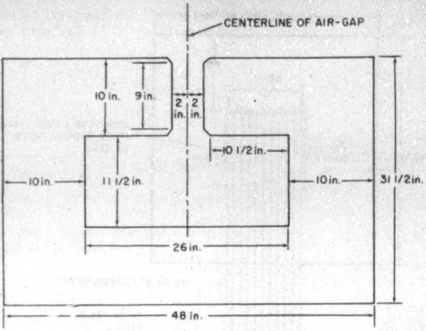


Fig. 4. View of yoke without coils and braking section

The effects of fringing and leakage were taken into account by empirical and analytical methods (see reference 6, and equation 14 of reference 7). The first of these, the fringing effect, causes an increase in the magnetomotive force as denoted by the fringing factor f_1 in the equation

$$AT_G = 0.313\ B_G d_2 f_{21} \tag{2}$$

From page 14 of reference 6, $f_1 = 1.5$; thus $AT_G = 0.313 \times 35{,}800 \times 4 \times 1.5 =$ 67,100 ampere-turns.

The second effect, the leakage factor f_2 was used in calculating the magnet yoke dimension d_6 in the equation

$$d_4 \times d_6 = \frac{f_2 B_G A_G}{B_m} \tag{3}$$

$$f_2 = 1 + \frac{3.4d_2(d_4+d_6)}{d_4 d_6} \times \left[\frac{d_8}{d_2+d_8} + \frac{0.45d_4}{d_2+0.67d_4+2d_8}\left(1+\frac{d_2+2d_8}{d_4}\right)\right] \tag{4}$$

(from reference 7).

The values of the quantities in the right-hand member of equation 4 are listed below. It will be noted that a value of 10 inches was assumed for d_6, which resulted in a square cross section of the yoke. The procedure was to calculate f_2, then solve for B_M in equation 3. If the value of B_M thus determined had been greater than the saturation flux density of 115,000 lines per square inch, it would have been necessary to assume a larger value for d_6. As shown by the calculations that follow, the value of 10 inches is satisfactory. In inches,

$d_2 = 4$
$d_4 = 10$
$d_6 = 10$
$d_8 = 0.5$

Then, $f_2 = 2.86$ and upon substituting this value in equation 3 and solving for B_M, it was found that $B_M = 102{,}000$ lines per square inch, which is a satisfactory operating flux density for the steel yoke. The reluctance drop in the steel from one side of the air gap, through the yoke around to the other side, was found to be 8,880 ampere turns.[8] This amount was added to the air-gap ampere-turns calculated in equation 2, to determine the wire size necessary for the coils.

Coil Design

The next step in designing the brake was to assume a coil current I of 10 amperes, direct current. Then, under the assumption that the coil is square in cross section, $d_4 = d_5 = 10$ inches.

The mean length of turn d_7 was calculated by the equation

$$d_7 = 2(d_4+d_5) + 2(d_6+d_5) = 2(d_4+2d_5+d_6)$$

$$d_7 = 80 \text{ inches}$$

The cross-section area of the coil conductor for a winding temperature of 60 C (degrees centigrade), or 140 degrees Fahrenheit, was calculated from the equation

$$A = \frac{kd_7NI}{1.16E} \tag{5}$$

(from reference 9).

where

$$k = \frac{100}{98} + 0.004(60-20) = 1.18 \text{ for copper of } 98\% \text{ conductivity at 60 C}$$

$$NI = \frac{67{,}100+8{,}880}{2} = 37{,}990 \quad \text{ampere-turns per coil}$$

$$E = 120 \text{ volts}$$

$$A = \frac{1.18 \times 80\ 37{,}990}{1.16 \times 120} = 25{,}200 \text{ circular mils}$$

No. 6 American Wire Gage double glass insulated wire was used for winding the coils, each of which was wound with approximately 3,400 turns. The cross section of each coil is approximately 10 by 10 inches; therefore the initial choice of magnet dimensions was correct.

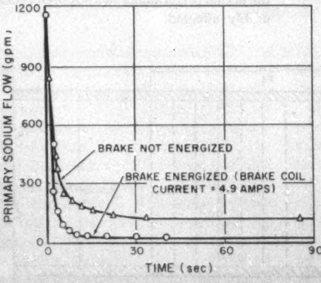


Fig. 5. Flow of sodium in SRE main primary piping circuit reactor to heat exchanger versus time, after reactor shutdown

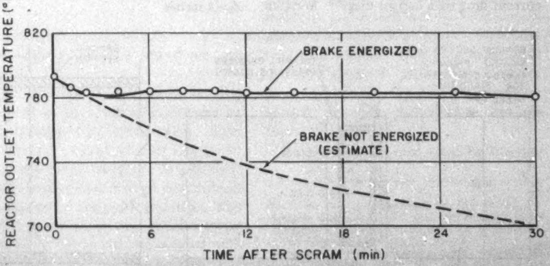


Fig. 6. Temperature of sodium at SRE reactor outlet to primary system versus time, after reactor shutdown

POWER LOSS IN COILS

The resistance of each coil at the operating temperature was calculated from

$$r_2=\frac{kd_7N}{1.16A} \qquad (6)$$

(from reference 9)

where

$$r_2=\frac{1.18\times 80\times 3{,}400}{1.16\times 22{,}700}=12.2 \text{ ohms}$$

Power loss in each coil = 10×120 = 1,200 watts.

Brake Performance

The effects of the eddy-current brake on the primary system are shown in Figs. 5 and 6. These curves illustrate the manner in which sodium flow in the primary system and sodium temperature at the reactor outlet vary with time for the condition where the brake is not energized, and the condition where the brake is energized, at the time of reactor shutdown.

It is possible to make the dragging pressure a maximum by variation of the dimensions of the brake, e.g., d_3, by taking $\partial P/\partial d_3$ from equation 1, setting $\partial P/\partial d_3=0$, and solving for d_3.

Nomenclature

P = dragging pressure in the sodium, psi
Q = sodium flow through the braking section, gallons per minute
d_1,d_2,d_3,d_4 = dimensions of braking section, shown in Figs. 1, 2, and 3, inches
B_G = flux density in sodium, lines per square inch
r_1 = resistivity of sodium, ohm-inches
f_1 = fringing factor (dimensionless)
f_2 = leakage factor (dimensionless)
AT_G = ampere-turns for air gap
B_m = flux density in steel yoke, lines per square inch
d_5 = coil dimensions shown in Fig. 1, inches
d_6 = yoke dimension shown in Fig. 1, inches
d_7 = mean length of coil turn, inches
d_8 = space between edge of coil and pole face, inches
A = cross-section area of coil conductor, circular mils
k = ratio of resistivity of copper conductor, used in coil winding to that of standard annealed copper at 20 C. For copper of $C\%$ conductivity at T C

$$k=\frac{100}{C}+0.004(T-20)$$

N = number of turns per coil
I = coil current, amperes
E = coil voltage
r_2 = resistance of each coil at operating temperature, ohms
A_G = area of pole face, square inches
$A_G=d_4\times d_6$

References

1. THE SODIUM GRAPHITE REACTOR ELECTRIC POWER PLANT, **S. Siegel.** *Electrical Engineering*, vol. 76, no. 1, Jan. 1957, p. 60.

2. ENERGIE DER WIRBELSTROME IN ELEKTRISCHEN BREMSEN UND DYNAMOMASCHINEN, **R. Rudenberg.** *Sammlung Elektrotechnischer Vortrage*, Stuttgart, Germany, vol. X, 1907, p. 269.

3. CALCULATION OF DEVELOPED PRESSURE AND FLUID POWER IN LINEAR POLYPHASE INDUCTION LIQUID METAL PUMPS, **R. S. Baker.** *Technical Report no. 48*, Mine Safety Appliances Company, Pittsburgh, Pa., Mar. 24, 1956.

4. MAGNETIC CIRCUITS AND TRANSFORMERS (book), **MIT Staff.** John Wiley & Sons, Inc., New York, N. Y., 1943, footnote p. 76.

5. RADIO INSTRUMENTS AND MEASUREMENTS. *Circular C74*, National Bureau of Standards, U. S. Government Printing Office, Washington, D. C., p. 294.

6. PERMANENT MAGNET DESIGN, **Earl M. Underhill.** Crucible Steel Company of America, Pittsburgh, Pa., 1957.

7. ESTIMATING LEAKAGE FACTORS FOR MAGNETIC CIRCUITS BY A SIMPLE METHOD, **R. K. Tenzer.** *Applied Magnetics*, Indiana Steel Products Company, Valparaiso, Ind., vol. 5, no. 2, Apr.–June 1957.

8. ELECTRICAL STEEL SHEETS. *Engineering Manual*, U. S. Steel Corporation, Pittsburgh, Pa., fourth edition, 1955, p. 258.

9. ELECTRICAL ENGINEERS' HANDBOOK (book), **H. Pender, W. Del Mar.** John Wiley & Sons, Inc., fourth edition, 1949, section 6–31.

Difference Equations and Their Applications

LOUIS A. PIPES
MEMBER AIEE

Synopsis: This paper is intended to introduce the methods and literature of difference equations to engineers who are unaquainted with them or their wide applicability and power.

The nature of the calculus of finite difference is presented and the basic operators used in this calculus are defined. The general theory of linear difference equations with constant coefficients is discussed in detail and applied to the solution of typical electrical and mechanical problems. The paper is concluded with a short discussion of the use of difference equations for the approximate solution of ordinary and partial differential equations. A short summary of additional applications of difference equations is given with reference to the current literature.

DURING the last 15 years there has been a great deal of activity in the application of difference equations in the solution of many problems that arise in statistics, science, and engineering.

The development of high-speed digital computing machinery has motivated the use of difference equations as approximations to ordinary and partial differential equations.[1–3] The use of difference equations for solving partial differential equations was discussed in 1928 in a celebrated paper by Courant, Friedrichs, and Lewy.[4] However, the methods presented in this paper were not put into practical use until about 1943 as a result of the stimulus of war-time technology and with the assistance of the first digital automatic computers. Problems involving time-dependent fluid flows, neutron diffusion and transport, radiation flow, thermonuclear reactions, and problems involving the solution of several simultaneous partial differential equations are being solved by the use of difference equations throughout the country.

Besides the use of difference equations as approximations to ordinary and partial differential equations, they afford a powerful method for the analysis of electrical, mechanical, thermal, and other systems in which there is a recurrence of identical sections.

The study of the behavior of electric-wave filters, multistage amplifiers, magnetic amplifiers, insulator strings, continuous beams of equal span, crankshafts of multicylinder engines, acoustical filters, etc., is greatly facilitated by the use of difference equations. The usual methods for solving such systems are generally very lengthy when the number of elements involved is large. The use of difference equations greatly reduces the complexity and labor involved in the solution of problems of this type.

Nature of the Calculus of Finite Differences

As is well known, the most important concept of mathematical analysis is that

---

Paper 60-203, recommended by the AIEE Basic Sciences Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., January 31–February 5, 1960. Manuscript submitted September 15, 1958; made available for printing December 22, 1959.

LOUIS A. PIPES is with the University of California, Los Angeles, Calif.

of a function. If to a given value of x a certain value of y corresponds, it is said that y is a function of the independent variable x and one writes symbolically $y(x)$. Two types of functions occur in the applications of mathematical analysis to physical problems; first there are functions in which the variable x may take every possible value in a given interval, that is, the variable x is continuous. Functions of this type are studied in the ordinary Differential and Integral Calculus.

A second type of function exists in which the variable x takes only the given values $x_1, x_2, x_3, \ldots x_n$. In this case, the variable x is discontinuous. The methods of the ordinary calculus are not in general applicable to such functions. The calculus of finite differences deals especially with functions of the second type, but it may be applied to functions of the first type as well. The origin of the calculus of finite differences is usually ascribed to Brooks Taylor who published his, "Methodus Incrementorum" in 1717. He pointed out that the relations between the terms of arithmetic and geometric progressions, which were known for a long time, represented the simplest examples of difference equations.

There are many analogies between the differential and integral calculus and the calculus of finite differences. For example, in the calculus of finite differences, the first difference, which is denoted by $\Delta(h)F(x)$ is defined by,

$$\Delta(h)F(x)=F(x+h)-F(x) \qquad (1)$$

In the differential calculus, the first derivative of a function $F(x)$ is defined by the equation,

$$DF(x)=\operatorname*{Limit}_{h\to 0}\frac{[F(x+h)-F(x)]}{h}$$

$$=\operatorname*{Limit}_{h\to 0}\frac{[\Delta(h)F(x)]}{h} \qquad (2)$$

It can be shown that the nth derivative of $F(x)$ may be expressed in the form,

$$D^nF(x)=\operatorname*{Limit}_{h\to 0}\frac{[\Delta(h)F(x)]}{h} \qquad (3)$$

The operator $\Delta(h)$ is called the forward difference operator. When there cannot be any misunderstanding about the increment h, the difference operator $\Delta(h)$ is mere written as Δ.

BASIC OPERATORS USED IN THE CALCULUS OF FINITE DIFFERENCES

The following fundamental operators are used widely in the Calculus of Finite Differences.

1. The Displacement Operator E

$$EF(x)=F(x+h) \qquad (4)$$

2. The Forward Difference Operator Δ.

$$\Delta F(x)=F(x+h)-F(x) \qquad (5)$$

3. The Backward Difference Operator ∇.

$$\nabla F(x)=F(x)-F(x-h) \qquad (6)$$

4. The Central Difference Operator δ.

$$\delta F(x)=F(x+h/2)-F(x-h/2) \qquad (7)$$

5. The Averaging Operator μ.

$$\mu F(x)=\frac{1}{2}[F(x+h/2)+F(x-h/2)] \qquad (8)$$

6. The Symbolic Representation of Taylor's Expansion.

$$e^{hD}F(x)=F(x+h),\ e=\text{base of natural logarithms},\ D=\frac{d}{dx} \qquad (9)$$

EQUIVALENT OPERATORS

If the fundamental basic defitions given previously are used, the following equivalent operators may be easily deduced.

$$E=e^{hD} \qquad (10)$$

$$\Delta=(E-1) \qquad (11)$$

$$\nabla=(1-E^{-1}) \qquad (12)$$

$$\delta=E^{1/2}-E^{-1/2} \qquad (13)$$

$$\mu=\frac{1}{2}\delta=\frac{1}{2}(E^{1/2}+E^{-1/2}) \qquad (14)$$

It can be shown[5-8] that except for a few restrictions, these operators can be manipulated according to the laws of ordinary algebra, and many relations between the basic operators may be obtained.

For example, one may write,

$$D=\frac{\ln E}{h}=\ln(1+\Delta)/h \qquad (15)$$

by the use of equations 10 and 11. Similarly by the use of equations 10 and 12, one has

$$\nabla=(1-E^{-1})=(1-e^{-hD}) \qquad (16)$$

It is easy to show that the operators δ and μ may be expressed in the forms

$$\delta=2\sinh(hD/2) \qquad (17)$$

$$\mu=\cosh(hD/2) \qquad (18)$$

An interesting and useful exposition of the symbolic calculus of operators will be found in a paper by Bickley.[9] By the use of these operators, and the relations between them, several finite difference formulas for interpolation, differentiation, integration, scanning, etc., may be obtained. These formulas are of great importance in numerical analysis.

For example, if the logarithm in equation 15 is expanded in powers of Δ, one obtains,

$$D=\frac{\ln(1+\Delta)}{h}=\frac{1}{h}(\Delta-\Delta^2/2+\Delta^3/3-\Delta^4/4+\ldots) \qquad (19)$$

This expression may be used to compute the derivative of a function in terms of its forward differences.

THE SCANNING OPERATOR, $S(h)$

In the theory of scanning, if one is given a function $F(z)$, it is required to obtain the functions $\phi(x)$ defined by,

$$\phi(x)=\int_{(q-h/)2.}^{(x+h/2)}F(z)dz \qquad (20)$$

If equation 20 is differentiated with respect to x, the result is,

$$D\phi(x)=F(x+h/2)-F(x-h/2)=\delta F(x) \qquad (21)$$

Hence,

$$\phi(x)=\frac{\delta}{D}F(x)=S(h)F(x) \qquad (22)$$

where $S(h)$ is the scanning operator defined by,

$$S(h)=\frac{\delta}{D}=\frac{2\sinh(hD/2)}{D}=\frac{h\sinh(hD/2)}{(hD/2)} \qquad (23)$$

If $\phi(x)$ is known, and it is desired to determine $F(x)$, one may write,

$$F(x)=S^{-1}(h)\phi(x) \qquad (24)$$

In equation 24, $S^{-1}(h)$ is the inverse scanning operator.

Linear Difference Equations with Constant Coefficients

The most important difference equations that arise in applications to electrical circuits, dynamics, structural problems, etc., are linear difference equations with constant coefficients.

These equations are of the type,

$$(a_nE^n+a_{n-1}E^{n-1}+\ldots+a_1E+a_0)y(x)=V(x) \qquad (25)$$

In the usual applications of difference equations, $h=1$, so that the operator E is defined by,

$$E^ky(x)=y(x+k),\ k=0, 1, 2, 3, \ldots \qquad (26)$$

Difference equations are classified in the same manner as differential equations. Equation 25 is called a linear difference equation of order n with constant coefficients, provided that the coefficients a_k are constants. If $V(x)=0$, then the equation is said to be homogeneous; if not, it is said to be nonhomogeneous. Let,

$$P(E)=(a_nE^n+a_{n-1}E^{n-1}+\ldots+a_1E+a_0) \qquad (27)$$

The general solution of equation 25 may be written in the form,

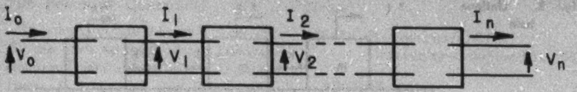


Fig. 1. Cascade connection of circuits

$$y=(y_c(x)+y_p(x)) \qquad (28)$$

If equation 28 is substituted into equation 27 the result is,

$$P(E)(y_c+y_p)=0 \qquad (29)$$

Now let,

$$P(E)y_c(x)=0 \qquad (30)$$

and

$$P(E)y_p(x)=V(x) \qquad (31)$$

By analogy to the solution of linear differential equations with constant coefficients, y_c is called the complementary function, and y_p is called the particular integral of the difference equation, equation 25.

DETERMINATION OF THE COMPLEMENTARY FUNCTION

To determine the complementary function $y_c(x)$, assume a tentative solution of the following form:

$$y_c(x)=Aq^x \qquad (32)$$

where A is an arbitrary constant and q is a number to be determined so that equation 32 is a solution of equation 30. It is noticed that if the operator E^k is applied to equation 32, the result is,

$$E^k y_c(x)=E^k Aq^x=q^k Aq^x=q^k y_c(x) \qquad (33)$$

It is therefore evident that if equation 32 is substituted into equation 30 the result may be written in the following form,

$$P(E)y_c(x)=P(q)y_c(x)=0 \qquad (34)$$

If the trivial solution $y_c(x)$ is excluded, then one must have,

$$P(q)=0 \qquad (35)$$

This is the characteristic equation of the difference equation, equation 25. The characteristic equation is an algebraic equation of the nth degree in q. The roots of this equation determine the possible values of q that allow equation 32 to be a solution of equation 30.

If the roots of equation 35 are the distinct numbers $q_1, q_2, \ldots q_n$, then the general solution of equation 30 may be written in the form,

$$y_c(x)=A_1 q^x{}_1+A_2 q_2{}^x+\ldots+A_n q_n{}^x \qquad (36)$$

The quantities A_k in equation 36 are arbitrary constants to be determined when the boundary conditions of the difference equation are given. The general solution of equation 36 is valid provided that the roots, q_k of the characteristic equation 35 are real or complex numbers provided that they are distinct.

If the characteristic equation 35 has a multiple root q_1 of the rth order, it can be shown[7] that the part of the general solution that corresponds to this rth fold repeated root is,

$$y_1(x)=(A_1+A_2x+\ldots+A_r x^{r-1})q_1{}^x \qquad (37)$$

where the A_k's are arbitrary constants.

THE PARTICULAR INTEGRAL

The difference equations most frequently encountered in practice are linear and homogeneous with constant coefficients. The various methods for obtaining the particular integrals of linear equations with constant coefficients have their counterpart in the theory of linear differential equations.[5,7,8] The special method to be used depends on the nature of the function $V(x)$. To illustrate the general procedure, let the function $V(x)$ be of the exponential form $V(x)=K\exp(mx)$. Let.

$$P(E)=(a_nE^n+a_{n-1}E^{n-1}+\ldots+a_1E+a_0) \qquad (38)$$

The difference equation for the particular integral is now,

$$P(E)y_p(x)=Ke^{mx} \qquad (39)$$

In this case, to obtain the particular integral assume a solution of the form,

$$y_p(x)=Ce^{mx} \qquad (40)$$

where C is to be determined. Now since,

$$E^k Ce^{mx}=Ce^{m(x+k)}=Ae^{mk}e^{mx} \qquad (41)$$

It follows that,

$$P(E)Ce^{mx}=Ce^{mx}P(e^m) \qquad (42)$$

Hence if equation 40 is substituted into equation 39, the following result is obtained:

$$Ce^{mx}P(e^m)=Ke^{mx} \qquad (43)$$

If $P(e^m)\neq 0$, then C is given by

$$C=K/P(e^m) \qquad (44)$$

the particular integral in this case is then,

$$y_p(x)=Ke^{mx}/P(e^m) \qquad (44A)$$

In general, the particular integral which is the solution of the equation

$$P(E)y_p(x)=V(x) \qquad (45)$$

may be obtained by an operational manipulation of equation 45 and can be expressed symbolically in the form,

$$y_p=P^{-1}(E)V(x) \qquad (46)$$

If $V(x)$ is a rational and integral function of x it is possible to expand $P^{-1}(E)$ in a series of ascending powers of Δ by the substitution $E=(1+\Delta)$, and operate with this series on $V(x)$. If $V(x)$ is a rational and integral function of x, the result of this operation will give a finite number of terms.

In some cases it is convenient to expand the operator $P^{-1}(E)$ into partial fractions. If this is done, the required particular integral may be obtained in terms of indefinite sums.[7]

Applications of Difference Equations to the Analysis of a Uniform Electric Network

An interesting and typical technical problem that illustrates the use of difference equations, is to determine the current and potential distribution of the electric circuit shown in Fig. 1, in the periodic steady alternating state.

The circuit of Fig. 1 is a cascade connection of n-boxes. Each box contains the Pi network of impedances Z and admittances $Y/2$ as shown in Fig. 2.

Each box of the network of Fig. 1 contains the Pi network of Fig. 2. This consists of a symmetrical arrangement of two shunt admittances $Y/2$ and a series impedance Z.

If the sending-end potential V_0 and the sending-end current are specified, it is required to determine the potential and current distribution of the uniform network. If Kirchhoff's first and second laws are applied to the network of Fig. 2, the following simultaneous difference equations are obtained:

$$I_{k-1}=I_k+\frac{1}{2}Y(V_k+V_{k-1}) \qquad (47)$$

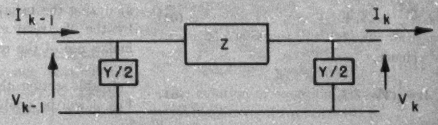


Fig. 2. Fundamental circuit

$$V_{k-1}=V_k+Z\left(I_{k-1}-\frac{Y}{2}V_{k-1}\right) \qquad (48)$$

In order to solve the simultaneous system of difference equations 47 and 48, it is necessary to separate the variables.

If equation 48 is solved for I_{k-1}, the result is

$$I_{k-1}=\frac{1}{Z}(V_{k-1}-V_k)+\frac{Y}{2}V_{k-1} \qquad (49)$$

If this equation is used to eliminate I_{k-1} and I_k in equation 47, the result is

$$V_{k+1}-(2+ZY)V_k+V_{k-1}=0 \qquad (50)$$

Equation 50 is a homogeneous difference equation of the second order of the type discussed in the previous section. Its solution is of the form,

$$V_k=Aq^k \qquad (51)$$

where A is an arbitrary constant and q satisfies the characteristic equation,

$$q^2-(2+ZY)q+1=0 \qquad (52)$$

The two roots of this equation satisfy the following equations:

$$q_1\cdot q_2=1 \qquad (53)$$

$$q_1+q_2=(2+ZY) \qquad (54)$$

Now let,

$$q_1=e^a, \quad q_2=e^{-a} \qquad (55)$$

where e is the base of the natural logarithms. Then,

$$(q_1+q_2)=(2+ZY)=2\cosh(a) \qquad (56)$$

THE PROPAGATION CONSTANT OF THE NETWORK

$$a=\cosh^{-1}(1+ZY/2)=\text{propagation constant} \qquad (57)$$

$$V_k=A_1e^{ak}+A_2e^{-ak} \qquad (58)$$

$$I_k=\frac{1}{Z}(V_k-V_{k+1})+\frac{Y}{2}V_k=\frac{1}{Z_0}\times[-A_1e^{ak}+A_2e^{-ak}] \qquad (59)$$

The quantity Z_0 is the characteristic impedance of the network and it is given by,

$$Z_0=\frac{2}{Y}\tanh(a/2) \qquad (60)$$

THE DETERMINATION OF THE ARBITRARY CONSTANTS

The two arbitrary constants A_1 and A_2 may be obtained from the following boundary conditions,

$$V_0=A_1+A_2 \qquad (61)$$

$$I_0=\frac{1}{Z_0}(-A_1+A_2) \qquad (62)$$

Hence,

$$A_1=\frac{1}{2}(V_0-Z_0I_0) \qquad (63)$$

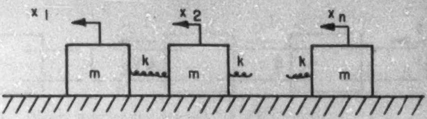


Fig. 3. Uniform train

$$A_2=(V_0+Z_0I_0)/2 \qquad (64)$$

If these values of A_1 and A_2 are substituted into equations 58 and 59 the results may be expressed in the following form,

$$V_k=V_0\cosh(ak)-Z_0I_0\sinh(ak) \qquad (65)$$

$$I_k=-V_0\frac{\sinh(ak)}{Z_0}+I_0\cosh(ak) \qquad (66)$$

where $k=0, 1, 2, \dots n$. To shorten the writing it is convenient to introduce the following matrix $[T]$. This matrix is known in the literature as the transmission or the chain-matrix of the network. This matrix is defined by the following equation,

$$[T]=\begin{bmatrix}\cosh(a) & Z_0\sinh(a)\\ \dfrac{\sinh(a)}{Z_0} & \cosh(a)\end{bmatrix}=\text{transmission matrix} \qquad (67)$$

If the following column matrix is introduced,

$$(y)_k=\begin{bmatrix}V_k\\ I_k\end{bmatrix} \qquad (68)$$

then equations 65 and 66 may be written in the compact matrix form,

$$(y)_k=[T]^{-k}(y)_0 \qquad (69)$$

From this it follows that one can write,

$$(y)_0=[T]^n(y)_n, \quad (y)_n=[T]^{-n}(y)_0 \qquad (70)$$

Longitudinal Oscillations of a Long Uniform Train

As another typical example of the use of difference equations in the solution of dynamical problems let it be required to determine the resonant frequencies and modes of oscillation of a chain of n cars that are coupled together by identical springs that satisfy Hooke's law. The system under consideration is depicted by Fig. 3.

Fig. 3 represents an idealized train of n cars of equal mass m couled to each other by identical springs of spring constant k. The effects of friction will be neglected in the analysis. The simplest method of analyzing the behavior of this system is by the electrical analog method which makes use of the results of the previous section.[9]

In this section, the dynamic system of Fig. 3 will be analyzed by the use of a chain of differential equations and difference equations.

If Newton's second law of motion is applied to the rth car in the chain of cars of Fig. 3, the following differential equation is obtained:

$$mD^2x_r+k(x_r-x_{r-1})+k(x_r-x_{r+1})=0, \quad r=2, 3, \dots (n-1) \qquad (71)$$

where $D^2=d^2/dt^2$. The set of differential equations 71 hold for the values of r indicated, but not for $r=1$ or $r=n$, (the two ends of the train).

In order to study the nature of the possible oscillations of the system, assume that the displacements x_r are executing simple harmonic oscillations of the form,

$$x_r=y_r\sin(wt+\phi), \quad r=1, 2, 3, \dots n \qquad (72)$$

where y_r are amplitudes of oscillation of the several cars, w is an angular frequency to be determined, and ϕ is an arbitrary constant.

If equation 72 is substituted into equation 71 and the result divided by $\sin(wt+\theta)$, one obtains,

$$y_{r+1}-2\theta y_r+y_{r-1}=0 \qquad (73)$$

where the quantity θ is given by,

$$\theta=(1-w^2m/2k) \qquad (74)$$

To solve the difference equation 73, assume a solution of the form,

$$y_r=Aq^r \qquad (75)$$

If this solution is substituted into equation 73, the following characteristic equation is obtained:

$$q_r-2\theta q+1=0 \qquad (76)$$

Let the two roots of this equation be q_1 and q_2 and let

$$q_1=e^a, \quad q_2=e^{-a} \qquad (77)$$

From equation 76 one has

$$(q_1+q_2)=(e^a+e^{-a})=2\cosh(a)=2\theta \qquad (78)$$

Hence, the quantity a is given by

$$a=\cosh^{-1}(\theta)=\cosh^{-1}(1-w^2m/2k) \qquad (79)$$

The solution of the difference equation 73 may therefore be written in the following convenient form:

$$y_r=B_1\cosh(ar)+B_2\sinh(ar), \quad r=2, 3, 4, \dots (n-1) \qquad (80)$$

where B_1 and B_2 are arbitrary constants.

Vol
Pt.
196

THE BOUNDARY CONDITIONS

In order to obtain the constants B_1 and B_2, use must be made of the boundary conditions at the front and rear of the train. It is realized that there are no cars corresponding to $r=0$ and to $r=(n+1)$. In order to extend the solution of equation 79 to take into account $r=0$ and $r=(n+1)$, let it be assumed that there are phantom cars that correspond to these indices. Since the front and rear of the train is able to move freely, there cannot be any strain in the phantom springs connecting the cr $r=0$ to the car $r=1$ or in the spring that is supposed to connect the car $r=n$ to the car $r=(n+1)$. This condition may be expressed mathematically by requiring that,

$$y_0=y_1 \text{ and } y_n=y_{n+1} \qquad (81)$$

Equation 80 can be used as if it were valid for $r=0$ and $r=(n+1)$.

The boundary conditions of equation 81 when expressed by equation 80, lead to the following two equations:

$$B_1=B_1 \cosh (a)+B_2 \sinh (a) \qquad (82)$$

$$B_1 \cosh (an)+B_2 \sinh (an)=B_1 \cosh [a(n+1)]+B_2 \sinh [a(n+1)] \qquad (83)$$

B_2 can be expressed in terms of B_1 by equation 82 in the form,

$$B_2=-\tanh (a/2)B_1 \qquad (84)$$

If this expression for B_2 is substituted into equation 83 and some algebraic reductions are made, the result is

$$B_1 \tanh (a/2) \sinh (an)=0 \qquad (85)$$

For a nontrivial solution, one must have $B_1 \pm 0$. Hence the possible values of the parameter a must satisfy the following transcendental equation:

$$\tanh (a/2) \sinh (an)=0 \qquad (86)$$

This equation is satisfied provided that,

$$\sinh (an)=0, \qquad (87)$$

This gives the following values for a,

$$a=a_s=j\pi s/n, \quad s=0, 1, 2, 3, \ldots (n-1), \quad j=(-1)^{1/2} \qquad (88)$$

If s is given the value $s=n$, equaton 86 leads to an indeterminate form and it is not satisfied.

THE RESONANT OR NATURAL FREQUENCIES OF THE TRAIN $=w_s$

To determine the possible natural angular frequencies of the system, one may use equation 78 written in the following form:

$$\cosh (a)=\cosh (a_s)=\theta=(1-w_s^2m/2k)=\cosh (j\pi s/n) \qquad (89)$$

Hence,

$$w_s^2m/2k=1-\cosh (j\pi s/n)=1-\cos (\pi s/n)=2 \sin^2 (\pi s/2n) \qquad (90)$$

Therefore,

$$w_s=2(k/m)^{1/2} \sin (\pi s/2n), \quad s=0, 1, 2, 3, \ldots (n-1) \qquad (91)$$

These are the n natural angular frequencies of the principal modes of oscillation of the train. The frequency $w_0=0$, corresponds to a rigid body translation of the entire train.

THE NATURAL MODES OF THE OSCILLATION OF THE TRAIN $=y_{rs}$

To determine the natural modes of the oscillation of the system write equation 80 in the form,

$$y_r=B_1 \cosh (ar)+B_2 \sinh (ar) \qquad (92)$$
$$=B_1 [\cosh (ar)-\tanh (a/2) \sinh (ar)], \quad r=1, 2, 3, \ldots n$$

To each value of $a=a_s=j\pi s/n$, there corresponds an amplitude function,

$$y_{rs}=B_s\phi(r, s) \qquad (93)$$

where

$$\phi(r, s)=[\cos (\pi rs/n)-\tan (\pi s/2n) \sin (\pi rs/n)] \qquad (94)$$

$r=1, 2, 3, \ldots n$ (cars)

$s=0, 1, 2, 3, \ldots (n-1)$ (modes)

The general solution of the chain of differential equations 71 with the appropriate boundary conditions can now be obtained by substituting equation 93 into equation 72 and summing over all the possible modes s.

The general solution of the chain of difference-differential equations 71 is, therefore,

$$x_r=\sum_{s=0}^{s=(n-1)} B_r\left[\cos\left(\frac{\pi rs}{n}\right)-\tan\left(\frac{\pi s}{2n}\right)\times \sin\left(\frac{\pi rs}{n}\right)\right] \sin (w_st+\theta_s) \qquad (95)$$

The index r refers to the car number and the index s pertains to the mode. The quantities B_r and θ_s are the $2n$ arbitrary constants of the general solution. These constants may be obtained if the initial displacements $x_r(0)$ and the initial velocities $\dot{x}_r(0)$ of each car is given.

Difference Equations Method for Solving Ordinary and Partial Differential Equations

In many problems of practical importance it is frequently not possible to solve the resulting ordinary or partial differential equations by known methods. However, it is sometimes possible to solve certain difference equations whose solutions approximate those of the original differential equations. In such cases one first approximates the differential equation by a difference equation and then solves the resulting difference equation. In this section the general procedure by which this is done is explained by means of typical examples.

From the definition of the derivative operator D and the operators Δ, ∇, and δ, it is easily seen that the following operators are approximations for the derivative operator D,

$$D \doteq \Delta/h \qquad (96)$$

$$D \doteq \nabla/h \qquad (97)$$

$$D \doteq \delta/h \qquad (98)$$

provided that the tabular interval h is sufficiently small. As $h \to 0$, the right members of these equations tend to D.

To illustrate the use of these finite difference approximations of the operator D, let it be required to obtain the finite difference approximation of the solution for the simple ordinary differential equation,

$$D^2y=x \qquad (99)$$

Let the initial conditions of equation 99 by $y(0)=0$ and $Dy(0)=1$. By the use of equation 98 and the operational equivalent, equation 13, the second derivative of the left member of equation 99 may be approximated by the following expression;

$$D^2y=(\delta/h)^2y=(E^{1/2}-E^{-1/2})^2y/h^2$$
$$=(E-2+E^{-1})y(x)/h^2 \qquad (100)$$

where h is small.

Hence the differential equation 99 may be replaced by the following approximate difference equation:

$$[y(x+h)-2y(x)+y(x-h)]/h^2=x \qquad (101)$$

The initial conditions can be written in the form,

$$y(0)=0, \quad [y(h)-y(0)]/h=1 \qquad (102)$$

Now let the difference equation 100 be required to be satisfied at the successive points $x=0, h, 2h, 3h, \ldots kh$. In such case the difference equation 100 may be written in the following form,

$$y_{k+1}-2y_k+y_{k-1}=kh^2, \quad k=1, 2, 3, \ldots \qquad (103)$$

where

$$y_k=y(kh) \qquad (104)$$

The initial conditions then take the form,

$$y_0=0, \quad y_1-y_0=h \qquad (105)$$

The solution of equation 103 with the initial conditions of equation 105 is

$$y_k = k(h-h^3)/6 + h^3k^3/6 \qquad (106)$$

The corresponds to the solution of equation 99 of the form,

$$y(x_k) = x_k + \frac{x^3}{6^k}k - \frac{h^2}{6}x_k \qquad (107)$$

The exact solution is,

$$y(x) = x + \frac{x^3}{6} \qquad (108)$$

In this case if h is so small that h^2 can be neglected, equation 107 tends to equation 108.

PARTIAL DIFFERENTIAL EQUATIONS

One of the most powerful methods for solving partial differential equations numerically when analytic methods fail because of the complexity of the differential equations or the boundary conditions is to use the method of finite differences.

The procedure is similar to the one used for the approximate solution of the ordinary differential equation 99. The differential equation is first approximated by a difference equation and then this resulting difference equation is solved subject to the appropriate boundary and initial conditions of the problem.

The partial differential equations that arise in the solution of physical problems are usually of the following form:

$$AU_{xx} + 2HU_{xy} + BU_{yy} + 2GU_{xz} + 2FU_{yz} + CU_{zz} = I \qquad (109)$$

In this equation the quantitites A, B, C, F, G, H, and I may be constants (including zero) or functons of (x,y,z). The dependent variable is $U(x,y,z)$ and the independent variables are (x,y,z). The notation U_{xx} denotes the second partial derivative of U with respect to x, etc. For example, if $A=B=C=1$ and $H=G=F=0$, then equation 99 reduces to the Poisson equation,

$$U_{xx} + U_{yy} + U_{zz} = I(x, y, z) \qquad (110)$$

If U is a function of x and y only, then equation 99 reduces to a 2-dimensional partial differential equation. This equation then takes the form:

$$AU_{xx} + 2HU_{xy} + BU_{yy} = I(x, y) \qquad (111)$$

In the standard literature on partial differential equations, equation 110 is classified into the following forms, depending on the nature of the coefficients A, B, and H.

1. Hyperbolic Equation.[10]

If $H^2 - AB > 0$, equation 111 is said to be hyperbolic.

2. Parabolic.[10]

If $H^2 - AB = 0$, equation 111 is said to be parabolic.

3. Elliptic.[10]

If $H^2 - AB < 0$, equation 111 is said to be elliptic.

THE POISSON EQUATION IN TWO DIMENSIONS

The 2-dimensional Poisson differential equation will be taken as an example of the method of finite differences to solve a partial differential equation. In this case the Poisson equation 110 takes the following form:

$$(D_x^2 + D_y^2)U(x, y) = I(x, y) \qquad (112)$$

In this case one may use the same approximation that was employed in equation 100 and write,

$$D^2_x = (E_x - 2 + E_x^{-1})/h^2 \qquad (113)$$

$$D_y^2 = (E_y - 2 + E_y^{-1})/h^2 \qquad (114)$$

where the displacement operators E_x and E_y are defined by the following equations:

$$E_xU(x, y) = U(x+h, y), \quad E_yU(x, y) = U(x, y+h) \qquad (115)$$

By the use of the operators in equations 113 and 114 the partial differential equation 112 may be written in the following approximate form,

$$U(x+h, y) - 4U(x, y) + U(x-h, y) + U(x, y+h) + U(x, y-h) = h^2I(x, y) \qquad (116)$$

A more convenient manner of writing equatin 116 is to express it in terms of the displacement operators E_x and E_y in the form:

$$(E_x + E_y + E_x^{-1} + E_y^{-1} - 4)U(x, y) = h^2I(x, y) \qquad (117)$$

A very ingenious method for the solution of the difference equation 117 in the interior of a rectangular domain has been given by P. C. Rosenbloom, in the paper, "Difference Equation Method for Solving the Dirichlet Problem."[11]

The classical example of the solution of a parabolic partial differential equation that represents the one-dimensional flow of heat in a conducting wire or the diffusion of a liquid or gas along a porous tube leads to the solution of the equation,

$$c^2U_{xx} = U_t \qquad (118)$$

with the initial conditions $U = f(x)$ at $t = 0$ and the two boundary conditions of the general type:

$$U + aU_x = b(t), \text{ at } x = x_1 \text{ and } x = x_2 \qquad (119)$$

This equation was studied by Courant, Friedrichs, and Lewy (see reference 4). It was discovered by these investigators that in order for the resulting difference equation 118 to have a solution that converges, it is necessary for the tabular intervals of the space and time variables h_x and h_t to be chosen in such a manner that the following inequality is satisfied:

$$c^2h_t/h_x^2 < 1/2 \qquad (120)$$

The relation 120 indicates that an unlimited decrease in the tabular interval h_x will not lead to improved accuracy of the solution unless it is accompanied by a suitable decrease in the tabular interval h_t.

Summary of Additional Applications

In this brief discussion of the applications of difference equations it is not possible to describe all the fields of applied mathematics in which they have been found useful. In conclusion, it may be mentioned that difference equations have been used in the following important types of problems:

1. The study of intermittent control systems.[12]
2. The analysis of systems with hereditary characteristics.[13]
3. The solution of certain problems in Probability.[14]

Difference equations have also been used in the solution of mathematical models of economic systems and in mathematical models used in the biological sciences.

Conclusions

It is hoped that this short exposition of the use of difference equations will serve to introduce the reader to the methods and the literature of difference calculus and will motivate engineers who are unaquainted with this subject to learn more about it and to apply it in the solution of technical problems.

References

1. THE NUMERICAL SOLUTION OF TWO-POINT BOUNDARY PROBLEMS IN ORDINARY DIFFERENTIAL EQUATIONS (book), L. Fox. Oxford University Press, London, England, 1957.

2. DIFFERENCE METHODS FOR INITIAL VALUE PROBLEMS (book), R. D. Richtmyer. Interscience Publishers, Inc., New York, N. Y., 1957.

3. CONSTRUCTION AND APPLICATIONS OF CONFORMAL MAPS, E. F. Beckenbach. National Bureau of Standards, Government Printing Office, Washington, D.C., Applied Mathematics series no. 18, 1952.

4. UBER DIE PARTIELLEN DIFFERENZENGLEICHUNGEN DER MATHEMATISCHEN PHYSIK, R. Courant, K. O. Friedrichs, H. Lewy. *Mathematische Annalen*, Leipzig, Germany, vol. 100, 1928, p. 32.

5. APPLIED MATHEMATICS FOR ENGINEERS AND PHYSICISTS, (book) L. A. Pipes. McGraw-Hill Book Company, Inc. New York, N. Y., second edition, 1958.

6. INTERPOLATION (book), J. F. Steffensen. London, England, 1927.

7. Calculus of Finite Differences (book), C. Jordan. Chelsea Publishing Company, New York, N. Y., 1947.

8. Treatise on the Calculus of Finite Differences (book), G. Boole. G. E. Stechert and Company, New York, N. Y., 1946.

9. Analysis of the Longitudinal Motions of Trains by the Electrical Analog, L. A. Pipes. *Journal of Applied Physics*, New York, N. Y., vol. 13, no. 12, Dec. 1942, pp. 780–86.

10. Partial Differential Equation of Mathematical Physics, (book), A. G. Webster. Dover Publications, Inc., New York, N. Y., 1953.

11. See reference 3, pp. 231–37.

12. Automatic and Manual Control (book), A. Tustin. Academic Press, New York, N. Y., 1952.

13. The Theory of Linear Operators (book), H. T. Davis. The Principia Press, Bloomington, Ind., 1936.

14. Operational Methods in Applied Mathematics (book), H. S. Carslaw. Oxford University Press, second edition, 1947.

15. Methods of Applied Mathematics (book), F. B. Hildebrand. Prentice-Hall, Inc., Englewood Cliffs, N. J., 1952, chap. II.

16. Difference and Associated Operators with Some Applications, W. J. Bickley. *Journal of Mathematics and Physics*, Cambridge, Mass., vol. 27, 1948, pp. 183–92.

◆

Discussion

W. E. Meserve, and **H. C. Torng** (Cornell University, Ithaca, N. Y.): The discussers would like to congratulate Professor Pipes on his lucid presentation on the art of the calculus of finite differences; especially that on linear difference equations with constant coefficients.

It is our opinion that in treating certain technical problems, for example the study of sampled-data control systems, the engineer often has to deal with nonlinear difference equations or linear difference equations with variable coefficients. It seems that it is to this area that efforts should be directed.

In a recent paper,[1] an approach is presented to extend the principle of first approximation to the field of nonlinear difference equations.

The discussers would also like to point out that one of the analogies between the differential and integral calculus and the calculus of finite differences, the concept of orthogonality, has been successfully used to study certain nonlinear sampled-data control systems. In terms of a set of orthogonal functions in the continuous domain, a periodic function can be expanded into a series of an infinite number of terms. An exact solution for certain nonlinear differential equations is therefore very difficult if not impossible, to obtain; while, in terms of a set of orthogonal functions in the discrete sense a periodic sequence can be replaced by a series with a finite number of terms. This property facilities greatly the task of identifying the exact solution of certain types of nonlinear difference equations.[2,3]

References

1. The Study of Second-Order Non-linear Difference Equations Containing Small Parameters, H. C. Torng. *Journal*, Franklin Institute, Philadelphia, Pa., Feb. 1960.

2. The Analysis of Non-linear Sampled-Data Control Systems Containing a Relay, a Saturating, or a Backlash Element, H. C. Torng. *Ph. D. Thesis*, Cornell University, Ithaca, N. Y., Feb. 1960.

3. Determination of Periodic Modes in Relay Servomechanisms Employing Sampled-Data. H. C. Torng, W. E. Meserve. *Transactions*, Professional Group on Automatic Control, Institute of Radio Engineers, New York, N. Y., vol. PGAC, 1960.

Louis A. Pipes: The author wishes to thank Mr. Meserve and Mr. Torng for their kind remarks on the lucidity of his paper. When this expository paper was being written, the author was not aware of the existence of the excellent work of Mr. Torng and Mr. Meserve on the applications of the methods of finite differences to the solution of nonlinear difference equations. These gentlemen are to be commended on the use of the concept of orthogonality of certain classes of functions to the study of an important class of nonlinear sampled-data control systems. The author is most grateful to Mr. Meserve and Mr. Torng for calling his attention to this very recent work and for including references to this work in their discussion.

A Magnetic-Tape-to-Paper-Tape Converter

M. RINGER
NONMEMBER AIEE

L. MINTZER
NONMEMBER AIEE

WITHIN THE LAST DECADE digital computation has evolved from the scientific computer with simple input-output devices to the present-day electronic data processors with their myriad of peripheral devices. This development is continuing with the addition of facilities to provide intercommunication between the central processor and remote points. Such a device has been developed by Honeywell for use with the DATAmatic 1000 System.

DATAmatic 1000 Data Processing System

The DATAmatic 1000 is a large-scale general-purpose electronic data processing system. It is composed of integrated business equipment employing high-speed computer principles designed to expedite all phases of record keeping and accounting. The processing units are co-ordinated to perform as a complete, highly efficient electronic office with the ability to handle a great variety and large volume of business operations quickly. Included is the handling of all steps in such operations as accounting, billing, payroll, production control, sales analysis, maintenance of name and address lists, inventory control, engineering computation, and numerous other office procedures.

The model 1500 Magnetic-Tape-to-Paper-Tape Converter (MTPT), Fig. 1, has been added to the DATAmatic 1000 family of peripheral devices as the first step in linking the DATAmatic 1000 to remote points via standard communication channels. The MTPT converts information stored on standard DATAmatic 1000 magnetic tapes to standard 5, 6-, 7-, and 8-level punched paper tapes at a rate limited only by the electromechanical speeds of present tape-punching mechanisms.

Magnetic Tape

The magnetic tape used in the DATAmatic 1000 System is 3 inches wide and has 31 information channels which run longitudinally along the tape. A tape block contains two 52-bit (binary digit) words in each channel. Blocks designated "first half of tape" are interlaced with those of "second half of tape." In this way, since intermittent motion with interblock stops is a common condition, the dead space required for acceleration and deceleration is actually used as

---

Paper 59-1190, recommended by the AIEE Computing Devices Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Chicago, Ill., October 11–16, 1959. Manuscript submitted June 17, 1959; made available for printing December 18, 1959.

M. Ringer and L. Mintzer are with the Minneapolis-Honeywell Regulator Company, Newton Highlands, Mass.

The authors wish to acknowledge the work of Dr. J. J. Eachus, Mr. S. D. Harper, Dr. W. L. Gordon, and Mr. W. M. Carey, who are chiefly responsible for the conception and design of the MTPT, and for their comments and criticisms which have considerably influenced the preparation of this material.

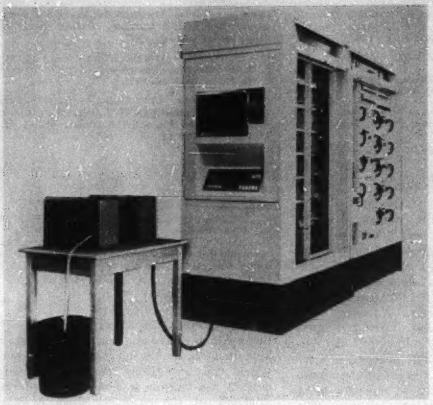


Fig. 1. Magnetic-tape-to-paper-tape converter

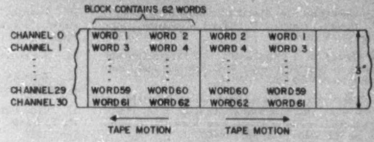


Fig. 2. Tape organization

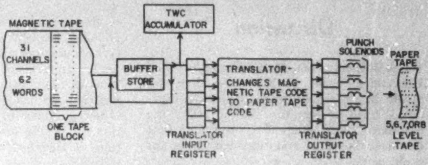


Fig. 3. Basic flow of information

active storage space for the other half of tape. Thus, high tape-storage efficiency obtained. One 2700-foot reel of DATAmatic 1000 tape stores the equivalent of 465,000 fully punched cards.

The tape moves at the rate of 100 inches or 100 blocks a second. A block is only 1/2 inch long but the foregoing rate prevails because of the interlacing feature. Acceleration and deceleration times are approximately 6 milliseconds. Thus it takes 18 milliseconds to start tape motion, read a block of information, and stop before the start of the next logical block.

A unique feature of the DATAmatic 1000 tape system is the ability to read tape moving in either direction. This considerably reduces access time, a desirable feature in any large file system.

Each of the 62 DATAmatic words in a magnetic tape block (Fig 2) consists of 48 bits of data plus four bits used for transfer checking. These four are called Transfer Weight Count (TWC) bits. In the MTPT the TWC bits are generated by making a mod 9 addition of the 48 data bits, bearing in mind their relative position in the word and weighting them accordingly. The mod 9 addition is discussed under the heating of "Logic." Consider the following example:

```
52 51 50 49 48 47 46 45................
                        9 8 7 6 5 4 3 2 1
TWC bits     8  4  2  1 ...............
                        1 8 4 2 1 8 4 2 1
```

Bit positions 1, 2, 3, and 4 have weights of 1, 2, 4, and 8 respectively. Similarly, bit positions 5, 6, 7, and 8 have weights of 1, 2, 4, and 8 respectively, etc. Whenever information is transferred from one part of the DATAmatic 1000 system to another, the TWC bits are recomputed and compared with those transferred with the word to check the validity of the information transferred.

MTPT Converter Used With DATAmatic 1000 System

The MTPT can be termed a hybrid unit in its use of components and circuits. It contains the magnetic core storage and logic circuits and the vacuum-tube control circuits that are standard with earlier converters; however, 75% of the converter contain newly developed transistorized circuits whose reliability has been established after many thousands of hours of error-free operation. These circuits, furthermore, are of such a form that they are intimately associated with Boolean expressions; hence one can derive from the logic equations all information necessary for their physical realization.

The MTPT is considered an output converter since it transfers information from magnetic tape to some terminal device such as, in this case, the paper-tape punch. Two earlier converters are the low-speed output converter which feeds a line printer at 150 lines per minute or a card punch at 80 cards per minute and the high-speed output converter which feeds a 900-line-per-minute printer.

All these converters provide the necessary buffer storage to accommodate the different instantaneous information rates of the magnetic tape and the particular terminal device. The converters, in addition, search for information on magnetic tape, select, edit, and translate from DATAmatic 1000 code to the codes required by the wide variety of output devices. Checking is an integral part of this process to assure error-free operation.

All these converters work with magnetic file (tape) units that operate via high-speed buffers with the central processor (CP). The converters are, however, "off line" devices, i.e., they function independently of the CP and hence require a magnetic file unit that has been released by the CP for such purposes.

The rate of processing information by the MTPT is 240 characters per second. Thus, it is obvious that for even the fastest paper tape punch, the MTPT is punch-speed limited and consequently only one channel of information buffer storage is required.

This paper will discuss the general system characteristics of the MTPT and the transistor circuits of the all solid-state portion of the equipment. In this portion, the variables of the logic equations are so intimately related to the components necessary to realize the logic that wiring information, component inventory etc., can be generated from the logic expressions alone. This will be described later.

BUFFERING

Fig. 3 illustrates the basic flow of information from magnetic tape to paper tape. The difference in the speed with which magnetic tape is read and the speed with which information is converted necessitates the use of a buffer stage. Information is read from the magnetic tape one channel at a time and stored in a buffer. It is possible to circulate this information in the buffer for the con-

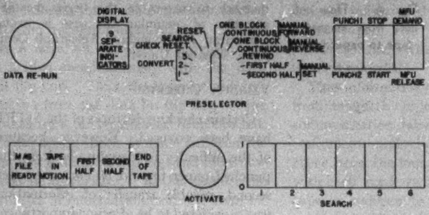


Fig. 4 (left). Control panel

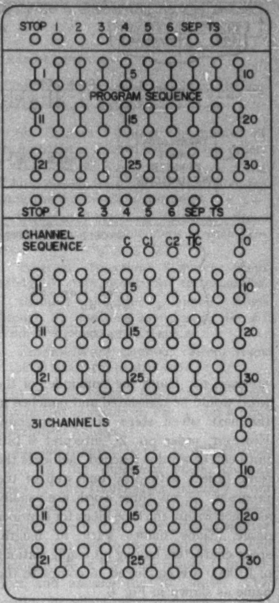


Fig. 5 (right). Program selection plug board

venience of conversion and for the purpose of TWC testing. The converter extracts information from the buffer one bit at a time. While conversion is taking place, the tape is mechanically repositioned in preparation for the next channel reading. The selection of the next channel is also accomplished during this period. Hence it is possible to utilize the time of conversion such that the rate at which the MTPT produces information is in no way limited by tape motion. The rate at which information is read into the buffers is controlled by the magnetic file unit. Once the information is in the buffer, the master clock of the MTPT takes over control of the rate at which the transistorized circuits perform. Finally, the punch determines the rate at which information is produced.

The buffer stage consists of two sections. One section, called "data store," is a 2-word store while the other section stores only one word and is called the "control word store." The control word serves two purposes. First, it indicates the position along the magnetic tape at which a particular conversion run is to commence. This is employed during the "search" mode of operation. Second, at the operator's convenience, it serves as an alternative to a plug board for the selection of programs.

DATA RETRIEVAL

The operator has the following facilities at his disposal for retrieving information and converting it into the desired form:

1. A rotary switch (Fig. 4) having the ability to select from 1 of 11 possible modes of operation.
2. Six toggle switches for establishing the search code.
3. Several self-explanatory push-buttons, lights, and toggle switches.
4. A program selection plug board (Fig. 5) which has the dual function of sequencing the 31 channels read into the buffer and sequencing the 6 available programs.
5. Program Plug Board. This plug board is divided into six major sections labeled Program 1–6. Each portion consists of 60 double hubs numbered 1–60 and performs the same function as can be handled by a sequence switch. These hubs are used in conjunction with the hubs positioned along the top and bottom of the panel. There are two sets of these hubs numbered from 1–22. The numbers represent the 22 conversion rules available in the converter which are listed in the Appendix. An impulse from a program hub to any of the conversion rule hubs will trigger the circuitry of the converter to execute the proper rule. The wiring between the program and conversion rule hubs determines the specific conversion rule to be executed during any given stage of the stepper. These 22 conversion rules can be plugged into 6 sections, each representing one program, using any desired valid format. It is therefore possible for a program to consist of 60 or less orders, made up of any combination of the 22 available orders. Generally, each program will be associated with a different format. For example, one program may convert the address of a message, one or more programs the content of the message itself, and yet another may be associated with the information concerning the source of the message.

With the aid of the foregoing facilities, the selection of data takes place in the following manner:

1. A block of information is selected, using part of the control word as a key. A comparison is made between the six search switch positions and the six high-order bits of the control word. A successful search results in the magnetic tape coming to a stop at the desired position.
2. Any one of the 31 channels in a block is selected via the program selection plug board.
3. Either of the two words in a channel can be selected by circulating the words in the buffer store. This is achieved by the use of conversion rule 19 which is "discard until end of word is reached."
4. The selection of bits within a word is accomplished by inserting the appropriate order in the desired position in a particular program.
5. Any one of the six available programs is selected by the program selection plug board, the control word or both.

TRANSLATION

A perusal of the Appendix will show that the order structure of the system is such that translation from the DATAmatic word to the desired tape code achieves the following results.

1. Alphanumeric, hexadecimal, and numeric interpretation.
2. Direct transcription.
3. Transcription of the TWC bits.
4. Generation of new information such as carriage return, dollar sign, decimal point, etc.
5. The omission of data.

The manner in which information is extracted from the store, translated, and punched on paper tape is best illustated by an example. Assume that the data store contains information, that the desired program has been selected, and that the first instruction of the program is "interpret 6 bits."

First, consider Fig. 6 which illustrates a 2-word channel on magnetic tape, in the data store and on punched paper tape. Order no. 1 transfers the first six bits from the data store into the translator input register. This is done one bit at a time, i.e., serially. When the six bits have loaded the input register, the translator fires and translation into a 5-level code (6-, 7-, or 8-level) is achieved in a single step. This results in the 5-level code for the letter *S* appearing in the translator output register causing the punch solenoids to be activated and the 5-level code *S* is punched on the paper tape. Order no. 1 could be repeated eight times in

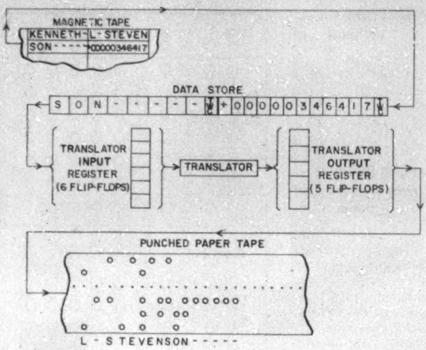

MAGNETIC TAPE
KENNETH-
L-STEVEN
SON----
0000346417
DATA STORE
S O N - - - - - + 0 0 0 0 0 3 4 6 4 1 7
TRANSLATOR INPUT REGISTER (6 FLIP-FLOPS)
TRANSLATOR
TRANSLATOR OUTPUT REGISTER (5 FLIP-FLOPS)
PUNCHED PAPER TAPE
L - S TEVENSON - - - - -

Fig. 6. Flow of data from magnetic tape to paper tape

program no. 1 and the remainder of the word would be translated and punched in the next seven steps of the program. However, order no. 2 "interpret 6 bits until end of word is reached" achieves the same object with only one step of the program. A full three words are usually assigned within the tape block for a proper name. Since names vary in length, space characters having bit configurations of 010000 are employed to complete the name as shown in Fig. 6.

The information bits are fed from the store serially into the translator input register and serially circulated in the store. The translator and punch are fed in parallel one character at a time.

With the completion of the first word, the next order in the program takes over and either interprets, transcribes, generates, or omits any number of bits desired.

A 52-step counter keeps track of the bits and their relative position, while a 60-step counter cycles through the program steps and yet another 30-step counter steps from program to program.

When a program has completed the orders required of it or when all but one of the program hubs have been used, then the last step will always be either "terminate program and consult the program sequence stepper" or "terminate program and consult the control word." The program stepper or control word will then direct the converter to the next program.

If desired, rule no. 9 which is "punch TWC bits" may be inserted at the end of each word. Any other form of check may be included in the 48 bits on magnetic tape and suitably interpreted and punched on paper tape. Any desired mode of information check may thus be included on the paper tape and so ensure the opportunity of validity checking at the receiving end.

Modes of Operation

The mechanical modes of operation indicated on the control panel shown in Fig. 4 are self explanatory. The three convert modes, however, merit some explanation. On teleprinters all letters of the alphabet are in lower case and all other symbols are in upper case. For each convert order, the MTPT determines whether upper or lower case is required. It further automatically inserts the required case symbol prior to punching the character which is being interpreted. At the beginning of a message, the MTPT inserts the required case symbol but thereafter remembers the previous case and only a case change effects the insertion of a required symbol.

The three modes of conversion are variations of the foregoing and are as follows:

1. The case memory is affected only by emission of an upper or lower case symbol (American Telephone and Telegraph Company).
2. The case memory is affected also by emission of a space thus causing the converter to act as if the remembered case is lower (Western Union).
3. If the remembered case was upper, then the emission of a space causes the case memory to revert to the begin message situation of assuming the case is wrong (American Telephone and Telegraph Company, Western Union).

Validity Checking

Certain checking features of the MTPT have been covered. However, because of the difficulty in visually checking the punched paper tape at 60 characters per second, reliable transfer of information was considered to be of prime importance, and hence many checking features have been incorporated in the device. The following paragraphs consider these features in detail.

TWC Checks. The data, having been loaded into the register, are circulated prior to processing for the purpose of TWC checking. If the test should reveal an error, the primary clock is interrupted, effectively preventing any further processing. The appropriate lamp is illuminated indicating the condition to the operator who then has the option of either recirculating the data and repeating the TWC check or processing the word in spite of the error. These two courses are effected by the operation of the data rerun and check reset positions of the selector switch. A successful TWC check permits the processing of the word. During processing, the TWC accumulator is appropriately augmented and a TWC check is made after processing each word. The latter test reflects any error that may occur in the register during processing.

The control word is TWC checked as it is loaded into the register, as is the data word and a similar indication is provided. If the control word is used during the processing of a block of information, it is TWC checked after that block has been processed. Again this gives assurance that no error has been generated in the control word register during processing.

Duplicate Data Bit Flip-Flop— Translator Output Register. As information is extracted from the data store, a data bit flip-flop is appropriately loaded and

Table I. Punch Check Sequence

Step No.	1		2	3		4		5
Action	Fire translator		Check that at least one of each pair is a "1"	Fire the solenoids and use the echo to reset the "1" side if the "0" side is a "0"		Reset the "0" side		Check that all flip flops are "0"
	"0" *F-F*	"1" *F-F*		"0" *F-F*	"1" *F-F*	"0" *F-F*	"1" *F-F*	
	1	0	If so, proceed to step 3. If not, stop clock and indicate	1	0	0	0	If check fails stop clock and indicate; otherwise proceed to next cycle
	0	1		0	0	0	0	
	0	1		0	0	0	0	
	1	0		1	0	0	0	
	1	0		1	0	0	0	

this information is subsequently loaded into the translator input register. A duplicate data bit flip-flop is provided and is loaded simultaneously with the data bit flip-flop. Once the transfer of information from the data bit flip-flop to the translator input register has been effected, a comparison check is made between the duplicate data bit flip-flop and the translator input register. In this manner the transfer of information from the data bit flip-flop to the input of the translator is checked.

Translator Output Register—Punch. Assuming the translator input register to be fully loaded, the next step is to fire the translator, the outputs of which appropriately set the translator output register. This register consists of two flip-flops per level, one representing a "zero" and the other a "one." Thus one and only one flip-flop of each level should be set at any one time. The first portion of the output register check is to establish that at least one flip-flop of each pair is set. This having been established, the solenoids of the punch are fired. The echos of the punch solenoids are then utilized to reset their associated "one" flip-flop of the output register if the zero member of that pair is a zero. Subsequently, the "zero" flip-flop of each pair is set to zero. The final check is then to establish that all flip-flops are zero. The foregoing process is shown in Table I. A 5-level output is assumed. Should the test fail, then the process error indication is activated and again the operator has the option of repeating or proceeding in spite of the error.

Program Checks. A check is provided which establishes that not more than one of the 22 available orders is activated at any one given time. Should this check fail, the program error indication is energized, the clock interrupted and the operator then takes appropriate action.

Another means of checking that the program is correctly assembled is provided. An order such as "Interpret 6 bits" may be inserted in error when less than six bits remain in that word. A check is provided which detects this type of program error and again energizes the program error indication.

Magnetic Tape Checks. Any error detected in the magnetic tape control circuitry stops the tape mechanism and energizes the magnetic file unit (MFU) error indicator. On reaching the physical end of tape, the MFU signals this condition by energizing the "tape physical end" indicator.

Paper Tape Run-Out Alarm. The paper tape carriage unit is adjusted so that a signal indicates the impending end of paper tape. The operator can then switch to the second punch and have the first reloaded.

Thus information is checked thoroughly from magnetic tape to punch coil. If communication takes place with distant equipment employing the identical TWC check, then the four TWC bits associated with each word are included on the paper tape in addition to the information. A "punch TWC bits" order exists for this purpose. By suitable programming any desired form of check bits can be included on the paper tape.

MAGNETIC SWITCHES

Liberal use is made of magnetic cores for realizing logic in areas where operations of a sequencing nature are desired. Also, the translator of the MTPT, i.e., the portion responsible for the translation from DATAmatic code to Teletype code, is a magnetic switch.

TRANSISTORIZED CIRCUITRY

The basic building block, i.e., the flip-flop, is illustrated in Fig. 7, it being contained within the broken-line rectangle. Significant features of this flip-flop are:

1. The use of a silicon diode in the base to provide an added 0.6-volt threshold in discriminating against noise.
2. The so-called K diodes which provide a second input to the flip-flop with the only component addition being a diode.

This second input is achieved by tying into the crossover gate. Note that this entry responds to a ground (0) signal whereas the base input responds to a negative (1) signal. Grounding K_1K_2 converts the flip-flop into a pair of amplifiers. In addition, the equations

$$\bar{A}=C \tag{1}$$

$$\bar{B}=A \tag{2}$$

are achieved by grounding one K. The K terminals appear on the connector and so make for considerable versatility merely by a connection on the connector.

BINARY COUNTER

This circuit is employed in achieving the following counters:

1. Six Counter. Counts the number of bits up to six associated with any one order.
2. Data Bit Counter. Keeps track of the 52 bits of each word and differentiates between the two words of a channel.
3. Command Selection Counter. Steps through the 60 possible steps of each program.
4. Program Selection Counter. Steps through 30 steps, sequencing the programs.
5. Channel Selection Counter. Steps through 31 positions representing the 31 channels from magnetic tape.
6. Control Word Counter. Keeps track of the 52 bits of the control word.

In all, 32 binary flip-flops are used. The circuit is shown in Fig. 7 and a brief description follows.

Assume A and C to be "1" and B to be "0". In the quiescent state, current flows in the transformer primary from transistor B via x, transformer, y, and register to -6 volts. Assume a negative pulse is applied to the input of C. The leading edge of the pulse causes C to conduct and assume the "0" state. This results in effectively applying a short circuit across the points x and y, i.e., across the primary of the transformer. The fall of current in the primary and the resulting transformer action causes transistor A to conduct. A becomes a "0" and by flip-flop action B a "1". The trailing edge of the pulse causes C to revert to the "1" state and the short circuit across x and y to be removed. This results in a rise of current through the primary of the transformer in the direction of y to x since the states of A and B are reversed. The back electromotive force caused by this rise in current is in the same direction as the back electromotive force due to the previous fall in current.

The electromotive force across the

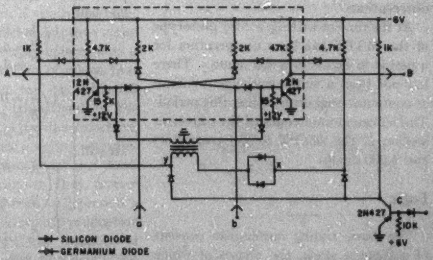


Fig. 7. Binary counter schematic diagram

secondary winding has the same polarity in both cases. Therefore, the leading edge of the input pulse switches the flip-flop and the trailing edge merely reiterates this state. The binary flip-flop, therefore, changes state once with each negative pulse.

PUNCH SOLENOID ENERGIZING CIRCUIT

This circuit, like the flip-flop package, achieves versatility of application by a suitable choice of connector wiring; see Fig. 8. Its basic function is to provide the drive for any circuit requiring current of the order of a quarter of an ampere to one ampere.

The input to this circuit is a standard logical function. The power transistor will be come active if either a "1" or a "0" is applied to the input of the circuit. The desired condition is established by a suitable choice of connector wiring.

CLOCK

Fig. 9 illustrates the basic circuitry of the master clock. The basic frequency of the clock is 50 kc. The square-wave output of each side is differentiated to give the *P* and *Q* pulses as illustrated in Fig. 10.

The timing elements of both the clock and differentiators are inductive. A paper[1] by W. M. Carey describes this type of circuit in detail. The function of the 51-ohm resistor is to bias the transistors slightly into the active region and so ensure self-starting. The function of the 1-kilohm resistor in series with the inductance is to ensure rapid recovery time.

Reliability and Margins of Operation

Included in the design specifications is the requirement that all circuitry be designed so that during a 25% variation of the supply voltages (−6, −13, and 6 volts) either separately or together, the MTPT is to operate reliably. The maintenance technician is required to vary the supply voltages accordingly during regular marginal tests and to maintain a record of his observations.

At the time of writing, a 20% prototype of the MTPT has been in operation for a period in excess of 8,000 hours. There has not been a single component failure or computational error during this period. The semiconductor count of the converter section of the MTPT is 850 transistors and 4,000 diodes.

Logic

The basic timing mechanism consists of a 2-phase clock, the phases of which are represented by *P* and *Q* pulses; see Fig. 10. The nominal frequency of the clock is 25 kc, resulting in a nominal separation of 20 microseconds between the *P* and *Q* pulses.

Logic is performed by two distinct types of circuitry: 1. flip-flops, amplifier, and diode gates, and 2. magnetic switches. The first type achieves a gate buffer form of logic and the magnetic switch realizes its "and" and "or" logic essentially by a suitable choice of wiring.

The subsequent discussion is concerned with the gate-buffer type of logic. Consider the equation

$$\bar{A} = CD + EFG \tag{3}$$

In the presentation of the logic, the following ground rule is established: The function to be generated is that effect which a number of inputs would have on an inverter. Furthermore, any one of the inputs is effective only when it is true and is represented on the right-hand side of the equality sign in the true state. Since an inverter realizes the complement of its inputs, the form in which it is presented is the complement of an affirmation.

In establishing the ground rules, it was decided that the collector of a conducting inverter would have the state of "0" and a nonconducting inverter the state of "1."

A flip-flop essentially consists of two inverters and two gates and is represented by the broken square enclosure in the following equations.

$$\bar{A} = B + CD \tag{4}$$
$$\bar{B} = A + E \tag{5}$$

The addition of inputs to the flip-flop is shown by a double input gate *CD* and a single input gate *E*.

The equations which follow describe the TWC accumulator as used in the MTPT:

2nd half

$$\bar{A}_3 = B_3C_3 + P\,B_1C_1 \qquad \bar{D}_3 = A_3 \tag{6}$$
$$\bar{B}_3 = A_3C_3 + P\,A_1C_1 \qquad \bar{E}_3 = B_3 \tag{7}$$
$$\bar{C}_3 = A_3B_3 + P\,A_1B_1 \qquad \bar{F}_3 = C_3 \tag{8}$$
$$\bar{A}_4 = B_4C_4 + P\,B_2C_2 \qquad \bar{D}_4 = A_4 \tag{9}$$
$$\bar{B}_4 = A_4C_4 + P\,A_2C_2 \qquad \bar{E}_4 = B_4 \tag{10}$$
$$\bar{C}_4 = A_4B_4 + P\,A_2B_2 \qquad \bar{F}_4 = C_4 \tag{11}$$

1st half

$$\bar{A}_1 = B_1C_1 + X_0F_3 + X_1F_3 + X_2F_3 + X_3E_3 + \text{Reset Function} \tag{12}$$
$$\bar{B}_1 = A_1C_1 + X_0D_3 + X_1F_3 + X_2D_3 + X_3F_3 \tag{13}$$
$$\bar{C}_1 = A_1B_1 + X_0E_3 + X_1D_3 + X_2E_3 + X_3D_3 \tag{14}$$
$$\bar{A}_2 = B_2C_2 + X_0C_3D_4 + X_0F_3F_4 + X_1A_3F_4 + X_1D_3D_4 + X_2C_3F_4 + X_2F_3E_4 + X_3A_3D_4 + X_3D_3E_4 + \text{Reset Function} \tag{15}$$
$$\bar{B}_2 = A_2C_2 + X_0C_3E_4 + X_0F_3D_4 + X_1A_3D_4 + X_1D_3E_4 + X_2C_3D_4 + X_2F_3F_4 + X_3A_3E_4 + X_3D_3F_4 \tag{16}$$
$$\bar{C}_2 = A_2B_2 + X_0C_3F_4 + X_0F_3E_4 + X_1A_3E_4 + X_1D_3F_4 + X_2C_3E_4 + X_2F_3D_4 + X_3A_3F_4 + X_3D_3D_4 \tag{17}$$

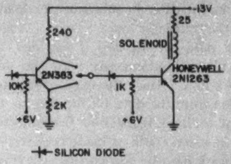


Fig. 8. Punch solenoid schematic diagram

The TWC accumulator consists of two pairs of registers, each pair having two tristable devices. A tristable device is recognized as those portions of the foregoing equations contained within the broken line squares. It comprises three amplifiers and three double-input gates.

Each of the four tristable devices is capable of assuming the three stable states indicated as follows, and no other stable states. All possible combinations of the stable states of the three inverters are thus exhausted.

A	*B*	*C*	*A*	*B*	*C*	*A*	*B*	*C*
0	1	1	1	0	1	1	1	0

The ternary digits are interpreted as follows:

0 1 1 as 0
1 0 1 as 1
1 1 0 as 2

A pair of tri-stable devices can thus assume the following nine distinct states:

A	B	State
0	0	0
0	1	1
0	2	2
1	0	3
1	1	4
1	2	5
2	0	6
2	1	7
2	2	8

The two halves of the TWC accumula-

Vol.
Pt. 1
196

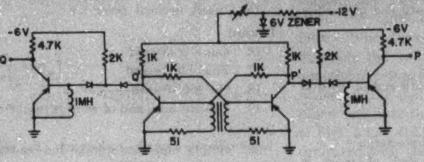


Fig. 9. Master clock schematic diagram

Waveforms at P, Q, P', and Q' are shown in Fig. 10

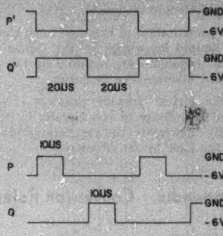


Fig. 10. Clock timing diagram

tor, each capable of nine distinct states, count modulo 9 as follows. Assume the accumulator to be in the "0" state. At the time of a Q pulse, a number is added to the first half. The succeeding P pulse results in this state being copied into the second half. The following Q pulse establishes a state in the first half which is the sum of the state of the second half and the new number of be added. This process continues as the weights of each of the 48 bits are added to the accumulator. Finally the state of the TWC accumulator is compared with the TWC bits of the word.

The functions $X0$, $X1$, $X2$, and $X3$ are dependent upon the Q pulse; bit position for weight; and knowledge as to whether the bit is a "1" or a "0". An $X2$ being a "one," for example, indicates that a bit having a weight of 4 is being added.

Consider Fig. 11 as representing the two registers of the accumulator and assume them both to be in the reset state. Now add 2 to the first half and transfer this state to the second half. The state of the first half at the next step will be the addition of the new number, say 8, and the state of the second half, which is 2. The new state of the first half is shown in Fig. 11.

The following chart indicates the new state of the first half subsequent to the addition of one of the X's to the second half:

State of Second Half	New State of First Half			
	X0	X1	X2	X3
00	01	02	11	22
01	02	10	12	00
02	10	11	20	01
10	11	12	21	02
11	12	20	22	10
12	20	21	00	11
20	21	22	01	12
21	22	00	02	20
22	00	01	10	21

Generation of Wiring Information

In the previous paragraphs it was observed that an exact correspondence exists between the logical equations and the hardware. It will now be shown how wiring information as well as maintenance and inventory records are generated directly from the equations. Distinct letters of the alphabet are allocated to represent different parts of the hardware. The letters A and B, associated as they are in the following equations, represent a flip-flop. Each of the letters D, E, F, K, R, V, and U represents a particular type of amplifier. The letters A, B, and C, associated as in the TWC accumulator, represent a tristable device. Figures added to the letters distinguish one of a kind from another of the same kind:

$$\begin{cases} \bar{A}_1 = B_1 + C_3 & (18) \\ \bar{B}_1 = A_1 + D_1A_4 & (19) \end{cases}$$

$$\bar{E} = A_1X_0 \qquad (20)$$

It is now possible to interpret the left-hand side of the equality sign in every equation in terms of the hardware. The right-hand side of the equations represents the inputs to amplifiers, flip-flops, etc., and are gates, the number of inputs being the number of letters per gate. The connection to a gate is indicated by the letter which states that it is directly connected to the output of the inverter represented by the letter. Thus in the foregoing equations, representing inputs to a flip-flop, the gate $D1A4$ has its inputs directly connected to the output of inverter D_1 and to the A_4 inverter of the A_4B_4 flip-flop. The output of the gate is directly connected to the b input of the A_1B_1 flip-flop. In addition, there exists standard flip-flop packages, amplifier packages, gate packages, etc. The input and output pin numbers of the packages are known.

It then becomes a fairly simple matter allocate a particular flip-flop to one of the three available positions in a package. Gates are allocated in a similar fashion. The foregoing information can thus make the listing of the interpin connections a simple matter. Therefore, it follows that sufficient information can be extracted from the equations, transferred to the input medium of a data handling machine, and effectively produce a wiring list automatically. This, in fact, has been done. The reader will recognize that certain aspects of the wiring may well present a difficult problem in automatic production of listings. However, the number of such problems was sufficiently small that a large percentage of the wiring information was automatically produced.

Conclusion

The gradual replacement of the human operator by the more efficient and speedy machine aids at the various stages of information transfer has received the attention of the development engineer since the beginning of communication systems. It is conceivable that an entire telegraphic system could take advantage of electronic computer techniques. Communication between an electronic computer and a distant point or in fact another computer remote from it, is a real need that is being met in varying degrees. Security, cost of communication link, and urgency are the factors that have determined the varying degrees with which a particular situation has been met.

The MTPT owes its development to the necessity to satisfy a need in this direction. The particular need was that in which a large quantity of information was to be retrieved from the magnetic tape file of a DATAmatic 10000 system and transmitted to a number of remote points. The action of retrieval and production of a fully edited paper tape was to be completely automatic. The function of the operator is restricted to that of activating a few switches and that of a monitor.

In developing the MTPT the following engineering objectives were maintained:

1. Use of machine aids wherever possible, in the production of wiring information.

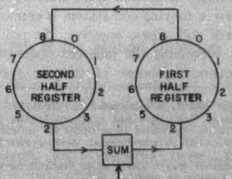


Fig. 11. Accumulator registers

2. Reliable and designable circuits.

3. A presentation of the logic such that a very limited knowledge of the logical designers techniques are required by the circuit and maintenance engineer.

4. The logical designer requires a very limited knowledge of the circuits in order to fully appreciate the restrictions placed upon the logic by the circuits.

Appendix. Conversion Rules

Alphanumeric
  1. Interpret 6 bits
  2. Interpret 6 bits until end of word is reached

Hexadecimal
  3. Interpret 4 bits
  4. Interpret 4 bits until end of word is reached

Numeric
  5. Interpret 4 bits
  6. Interpret 4 bits until end of word is reached

Transcription
  7. Transcribe one character
  8. Transcribe until end of word is reached

TWC
  9. Punch TWC bits

Emission
  10. Insert carriage return (CR)
  11. Insert line feed (LF)
  12. Inset space (SP)
  13. Insert blank (BL)
  14. Insert dollar sign ($)
  15. Insert decimal point (.)

Discard
  16. One bit discard
  17. Four bit discard
  18. Six bit discard
  19. Discard until end of word is reached

Control
  20. Empty tape date storage and read a new channel
  21. Terminate program and consult the control word
  22. Terminate program and consult the program sequence stepper

Reference

1. INDUCTIVELY CONTROLLED COMPUTER CIRCUITS, W. M. Carey. *Electronics*, New York, N.Y., Sept. 1959.

An Analysis of a Polarized Electromagnetic Drive System and of Vibrator Drive Systems

R. STEWART NUNN
NONMEMBER AIEE

Synopsis: Conventional electromagnetic contact-drive systems suffer from the disadvantage that resonant oscillation cannot be obtained, and this is shown in the analysis included. A scheme which overcomes this defect by the use of a permanent magnet and a tuned drive circuit is described. The drive power efficiency is shown to be greatly increased or alternatively for a given power input greater amplitudes of oscillation are obtained.

MECHANICAL OSCILLATION may be obtained electromagnetically either from an a-c or a d-c supply and, it has been shown in a previous paper by the author,[1] when the power input exceeds a certain minimum the drive force obtained when using a moving-armature is greater than that obtained when a moving-coil system is employed. There still arises, however, the problem of overcoming the inherent nonresonant frequency response of the d-c system. The problem is not difficult to overcome in a-c systems, provided a supply having the appropriate frequency is available, but in d-c systems resonant conditions cannot be obtained under normal circumstances, as will be seen subsequently.

By the use of a permanent magnet in the magnetic circuit of the actuating coil, and a tuned circuit also, energy that would normally be expended in opposing the motion is utilized in assisting it. Thus the power required to maintain a given oscillation is reduced. Alternatively, for a given power input larger amplitudes of oscillation are obtained.

The drive power has to supply the iron and copper losses of the actuating coil as well as the loss resulting from the internal molecular friction of the oscillating member and that resulting from windage. Also ohmic losses caused by charge and discharge of the shunt capacitor have to be accounted for in the conventional system (discussed later). The elimination of this and the reduction of the iron and copper losses at low oscillating frequencies will therefore enable systems to be operated at high frequency with improved efficiencies. The need for improved efficiency becomes greater as oscillation frequency requirements increase, and also when the dissipation of power in the contacts is appreciable in high power contact-drive systems.

Contact-Drive Systems

In electromagnetic contact-drive systems, such as that shown in Fig. 1, oscillation is obtained by means of an actuating coil energized from a d-c supply through contacts. One of these contacts is attached to the oscillating member which is attracted by the flux from the actuating coil when energized, so as to open the contacts. When this has occurred the potential energy continues to be stored in the compliance of the oscillating member throughout the period of its attractive excursion and thereby establishes its return, and while the energy is progressively converted into kinetic energy the contacts reclose. Repetition of events thus occurs, and during the dwell period of the contacts (which is achieved by mounting them resiliently) energy is imparted to the system, enabling continuous oscillation to be obtained from a constant voltage source.

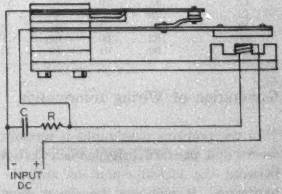


Fig. 1. D-c system to obtain continuous oscillation

Paper 60-201, recommended by the AIEE Basic Sciences Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., January 31–February 5, 1960. Manuscript submitted May 7, 1959; made available for printing January 26, 1960.

R. STEWART NUNN is with the Ministry of Aviation, London, England.

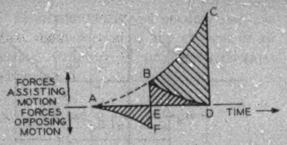


Fig. 2. Force-time diagram—conventional system

Unfortunately the derivation of the energy transferred to the mechanical system is not simply the total mechanical work done. During the first half of the closed period of the contacts it will be observed that the oscillating member is moving away from the force derived from the coil, and during the second half period it is moving towards the pole piece. Thus the net transference of energy to the system is the difference between the amounts of work done in each of these periods. That there is a balance of energy will be realized when the behavior of the coil inductance is taken into account. Not only does the coil flux establish the actuating force of the system, its time rate of change also resolves the gradual increase of force during the cycle, so that greater forces are manifested during the second half of the contact closed period than are manifested during the first. The force-time diagram in Fig. 2 will therefore indicate the unequal areas of half-contact-closed periods, area *BCD* representing the balance of energy necessary for oscillation.

If, therefore, the work done in opposing the motion of the oscillating member is reduced or eliminated there will be a greater transference of energy to the system, and improved drive-power efficiences will result.

This may be achieved by arranging the direction of the winding of the coil such that the flux emanating from it is opposed by the flux established by the permanent magnet, and therefore rendered inactive when the motion is against the coil force. When the motion is in the opposite direction the coil flux will be in the opposite direction (and therefore rendered active) if the coil shunt capacitance is tuned such that the waveforms as shown in Figs. 3 and 4(A) are obtained. Thus, compared with the conventional system a greater amplitude of oscillation is obtained, and this, moreover, is increased still further because of the fact that the energy that would normally be expended in opposing the motion is re-established [see Fig. 9(B)] such that it effectively boosts the active flux. It can be arranged that no energy such as that represented by area *AEF* (equal to area *DEB*) is expended, leaving the total coil energy (area *ACD*) available.

Fig. 3. Oscillograms of voltage waveform of tuned circuit

By referring to Figs. 4(A), (B), and (C) the sequence of events may be followed. If the coil resistance and iron losses are neglected, then when the drive contact is closed the induced voltage will be equal and opposite to the applied voltage. The rate of change of flux will be constant since the applied voltage is contant, so that the flux wave will be linear.

During the open period of the contacts the current (which is supplied by the coil shunt capacitance given by C^{dv}/dt) and the flux which is in phase with it will be zero at the reversed voltage peak *P*. When the voltage is zero the rate of change of flux will be zero since the voltage across the coil is given by $N^{d\phi}/dt \times 10^{-8}$. The flux will therefore reach a peak at these points, so that the waveform of flux will be an alternating one depending for its precise shape on the contact closure period. Thus an alternating wave of flux may be established from a d-c input by tuning the waveform of voltage, to that as shown in Figs. 3 and 4(A).

The alternating wave of flux is shown in Fig. 4(B) and it is seen that when the contacts close the flux derived from the coil is negative. It will also be observed that the oscillating member is moving away from the pole-piece, so that if the coil winding is in the direction such that the flux from it is opposed by the flux from the permanent magnet during the period *A* to *B*, the coil force opposing the motion may be reduced or eliminated. During the succeeding two quarter cycles, *B* to *C* and *C* to *D*, the flux derived from the coil is positive, so that it assists the motion (which is now towards the pole-piece) together with the reimbursed flux provided by the magnet. During the final quarter period *D* to *E* the flux is negative, but the oscillating member is moving away from the pole-piece, so that once again the permanent magnet neutralizes the coil-flux and it is not expended in opposing the motion.

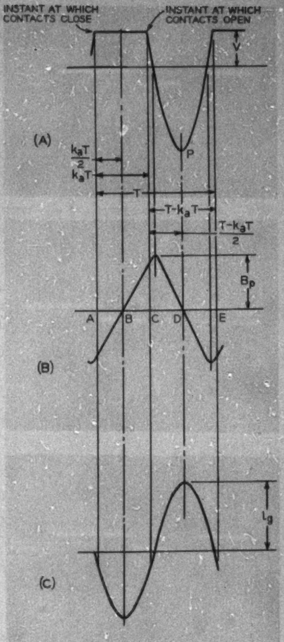


Fig. 4. Voltage, flux, and deflection diagrams

A—Voltage waveform
B—Flux waveform
C—Deflection

The effect of the permanent magnet therefore in an electromagnetic drive system using a tuned circuit, is to reverse energy derived from the coil that would be expended in opposing the motion to energy that is applied in assisting it. With the requisite strength of permanent magnet all the energy transferred to the mechanical system may be utilized to produce motion.

Analysis of the Motion Obtained

Consider first the conventional system in which the *C*, *R* leg in Fig. 1 is simply a spark suppression unit, and no permanent magnet is employed. The voltage waveform is as shown in the oscillograms Fig. 5. The flux waveform may then be taken to be as shown in Fig. 6(A); the

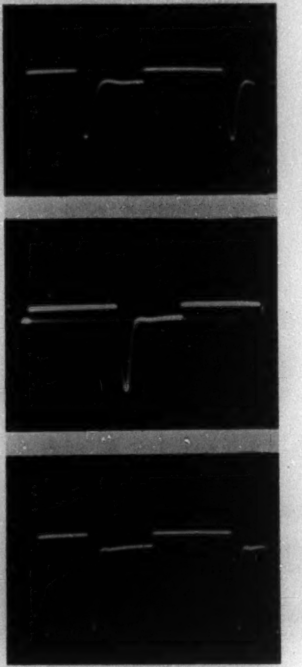


Fig. 5. **Oscillograms of voltage waveforms (various drive coils, conventional system)**

waveform of force will be as shown in Fig. 6(B), since the force F is given by

$$\frac{B^2A}{8\pi}\text{ dynes}$$

where

B = flux density in the air gap, gauss

A = cross-sectional area of the air gap, centimeter^2

Since the compliant and inertial forces set up by the drive contact on closure have little or no effect on the motion of the oscillating member, the waveform of force can be resolved into a series of harmonic waves, and accordingly the motion may be determined.

One has by Fourier's theorem

$$F=f(t)=a_0+\sum_{n=1}^{\infty}a_n\cos(nwt)+\sum_{n=1}^{\infty}\times b_n\sin(nwt)$$

and

$$F=\left\{\frac{2Bpt}{2\pi k_d}\right\}^2\frac{A}{8\pi}$$

where

$2B_p$ = the peak value of flux density attained

k_d = the closure period of the drive contact given by the ratio of

$$\frac{\text{time contact is closed per cycle}}{\text{time for the cycle}}$$

Since one wishes to examine the conditions for resonance by considering the phase displacement between force and deflection, reference will be made to peak values. The amplitude of oscillation occurs at the midpoint of the open and closed periods of the drive contact so that from Fig. 6(A) one obtains for $f(t)$

$$F=\left\{\frac{2Bp}{2\pi k_d}\right\}^2\{t+\pi k_d\}^2\frac{A}{8\pi}$$

The constant term

$$\frac{1}{2\pi}\int_0^{2\pi}f(t)dt$$

$$=\frac{Bp^2A}{16\pi^4k_d^2}\int_{-\pi k_d}^{\pi k_d}(t+\pi k_d)^2\,dt$$

$$=\frac{4Bp^2A}{3}\;k_d$$

The sine coefficients $\frac{1}{\pi}\int_{-\pi k_d}^{\pi k_d}f(t)\sin nt\,dt$

$$=\frac{Bp^2A}{8\pi^4k_d^2}\left\{\int_{-\pi k_d}^{k_d}(t+\pi k_d)^2\sin nt\,dt\right\}$$

$$=\frac{Bp^2A}{8\pi^4k_d^2}\left\{(2t+2\pi k_d)\frac{\sin nt}{n^2}-(t^2+2t\pi k_d+\pi k_d^2)\frac{\cos nt}{n}+\frac{2\cos nt}{n^3}\right\}_{-\pi k_d}^{\pi k_d}$$

$$=\frac{Bp^2A}{2\pi^3k_d}\left\{\frac{\sin[n\pi k_d]}{n^2}-\frac{\pi k_d\cos[n\pi k_d]}{n}\right\}$$

The cosine coefficients

$$\frac{1}{\pi}\int_{-\pi k_d}^{\pi k_d}f(t)\cos nt\,dt$$

$$=\frac{Bp^2A}{8\pi^4k_d^2}\left\{\int_{-\pi k_d}^{\pi k_d}(t+\pi k_d)^2\cos nt\,dt\right\}$$

$$=\frac{Bp^2A}{8\pi^4k_d^2}\left\{(t^2+2t\pi k_d+\pi^2k_d^2)\frac{\sin nt}{n}+(2t+2\pi k_d)\frac{\cos nt}{n^2}-\frac{2\sin nt}{n^3}\right\}_{-\pi k_d}^{\pi k_d}$$

$$=\frac{Bp^2A}{2\pi^3k_d}\left\{\frac{\pi k_d\sin[n\pi k_d]}{n}+\frac{\cos[n\pi k_d]}{n^2}-\frac{\sin[n\pi k_d]}{\pi k_d n^3}\right\}$$

As this is a rapidly converging series only the fundamental need be considered. Therefore the fundamental component of force producing oscillation will be

$$\frac{Bp^2A}{2\pi^3k_d}\left[\left\{\sin[\pi k_d]-\pi k_d\cos[\pi k_d]\right\}\sin nt+\left\{\left(\pi k_d-\frac{1}{\pi k_d}\right)\sin[\pi k_d]+\cos[\pi k_d]\right\}\cos\omega t\right]$$

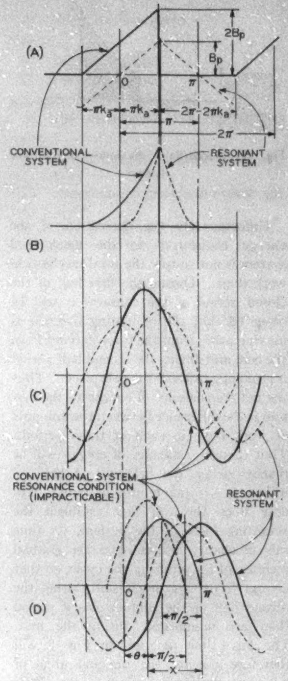


Fig. 6. Flux and force diagrams

A—Waveform of flux
B—Waveform of force
C—Fundamental harmonic components of force
D—Phase displacements between force and deflection

In order to be able to apply this to the determination of the motion obtained it is necessary to take into account the mass of the oscillating member as well as the mass of the armature. In the method first suggested by Lord Rayleigh and quoted by Timoschenko and Lessels[2] the kinetic energy of the oscillating member is determined by summation of the kinetic energy of every element of the member, from which a mass acting at the point of application of the drive force that is equivalent to the mass of the oscillating member is obtained. The mass of the member is normally small compared with the mass of the armature so that it makes little or no difference to the potential energy of deflection, but as stated by Timoschenko and Lessels the method is accurate even in cases where the mass of the oscillating member is equal to the mass of the armature.

From the theory of bending one has that the deflection of any element of an elastic structure of uniform section is given by

$$y=\frac{P}{EI}\left\{\frac{\text{le}\,x^2}{2}-\frac{x^3}{6}\right\}$$

where

x = any point along the length of the oscillating member measured from its root
le = total length from root
P = force acting
E = modulus of elasticity of the material
I = area moment of inertia of section

The deflection at the extremity of the elastic member is given by

$$\frac{P\,\text{le}^3}{3\,EI}$$

so that if V is the velocity of the armature at the point of application of the drive force (in this case at the extremity of the oscillating member) then the linear velocity of any element will be

$$\frac{P}{EI}\left\{\frac{\text{le}\,x^2}{2}-\frac{x^3}{6}\right\}\frac{V}{\dfrac{P\,\text{le}^3}{3EI}}$$

If q is the mass per unit length of the member its kinetic energy will be given by

$$\int_0^{\text{le}}\left\{\frac{3\,\text{le}\,x^2-x^3}{2\,\text{le}^3}\right\}^2\frac{V^2}{2}\,q\,dx$$

$$=\frac{33}{140}\,\text{le}\,q\,\frac{V^2}{2}$$

But le q is the total mass of the oscillating member, so that only 33/140 of it need be considered to be acting in addition to the mass of the armature. Let this mass equivalent and the mass of the armature equal M, then

$$M\frac{d^2y}{dt^2}+r\frac{dy}{dt}+Sy=\sqrt{H_s^2+H_c^2}\cos[wt-\theta]$$

where

r = viscous damping
S = stiffness of oscillating member
H_s = sine component of the fundamental harmonic of drive force
H_c = cosine component

The solution may be written

$$\frac{\sqrt{H_s^2+H_c^2}}{\sqrt{(S-\omega^2M)^2+r^2\omega^2}}\{\cos[\omega t-\theta-\psi]\}$$

to represent the motion after transient conditions have subsided

$$\psi=\tan^{-1}\frac{r}{\dfrac{S}{\omega}-\omega M}$$

and

$$\theta=\tan^{-1}\frac{H_s}{H_c}$$

Resonant conditions are obtained when the frequency of oscillation is equal to the undamped natural frequency, i.e.

$$\sqrt{\frac{S}{M}}=\omega \text{ and } \psi \text{ becomes 90 degrees}$$

As will be seen from Fig. 6(D) $\{\theta+\psi\}$ must always equal 180 degrees so that under conditions of resonance θ becomes 90 degrees also, so that H_c becomes zero, or

$$0=\{\pi^2k_d^2-1\}\sin[\pi k_d]+\pi k_d\cos[\pi k_d]$$

The solution is obtained when k_d is 0 and when k_d is 0.88 (approximately).

It is thus seen that under normal conditions resonant oscillation cannot be obtained in a conventional system. The relationship of phase angle with contact closure period is shown in Fig. 7.

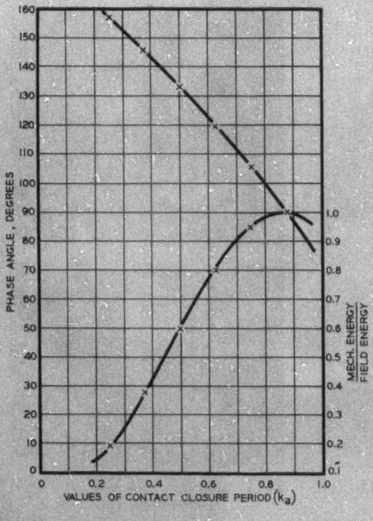


Fig. 7. Relationship of mechanical energy versus field energy and phase angle with various values of contact closure period

The system described in this paper can be analyzed in a similar manner. The waveform of force is shown in Fig. 6(B), the sine coefficient being

$$\frac{Bp^2A}{\pi^4}\left[\left\{\frac{1}{k_d}+\frac{1}{(1-k_d)}\right\}\pi\sin[\pi k_d]+\left\{\frac{1}{k_d^2}-\frac{1}{(1-k_d)^2}\right\}\cos[\pi k_d]-\frac{1}{k_d^2}-\frac{1}{(1-k_d)^2}\right]$$

and the cosine coefficient being

$$\frac{Bp^2A}{\pi^4}\left[\left\{\frac{1}{(1-k_d)^2}-\frac{1}{k_d^2}\right\}\sin[\pi k_d]+\left\{\frac{1}{k_d}+\frac{1}{(1-k_d)}\right\}\pi\cos[\pi k_d]\right]$$

When k_d is 1/2 the cosine coefficient is zero, so that resonant oscillation is obtained. If, however, the force is considered in its entirety, when k_d is 1/2 we almost have the classic case of a sinusoidal oscillation from a sinusoidal force. It will be observed that amplitudes greater than those obtained at resonance can be achieved, and moreover can be obtained at all practical values of contact closure periods whereas they can never be achieved at any value of closure period in the conventional system.

From the equation of motion of a sinusoidal system we obtain the amplitude of oscillation

$$=\frac{\text{Peak Harmonic Force}}{\sqrt{(S-\omega^2M)^2+\omega^2r^2}}$$

and w can be varied so as to obtain amplitudes greater than those obtained at resonance, so that

$$[(S-\omega^2M)^2+\omega^2r^2]$$

must be a minimum. This is when

$$\omega=\sqrt{\frac{S}{M}-\frac{r^2}{2M^2}}$$

Substituting this in the equation for amplitude of oscillation we obtain a maximum value attainable by the use of the resonant system:

$$=\frac{\text{Peak Harmonic Force}}{\sqrt{\left\{S-M\left(\frac{S}{M}-\frac{r^2}{2M^2}\right)\right\}^2+r^2\left\{\frac{S}{M}-\frac{r^2}{2M^2}\right\}}}$$

$$=\frac{\text{Peak Harmonic Force}}{2\pi\, r \text{ x True Natural Frequency}}$$

Energy Considerations

From Figs. (6A) and (B)

Actuating force throughout cycle=

$$\frac{A}{8\pi}\left\{\frac{2Bp}{2\pi k_d}\right\}^2\{t+\pi k_d\}^2$$

or

$$\frac{A}{8\pi}\left\{\frac{2Bp}{\frac{k_d}{f}}\right\}^2\left\{t+\frac{k_d}{2_f}\right\}^2$$

$$=\frac{Bp^2f^2A}{2\ \pi k_d^2}\left\{t+\frac{k_d}{2_f}\right\}^2$$

Assume sinusoidal oscillation is obtained, then

$$\text{deflection}=-\text{lg}_c\cos[\omega t]$$

Then the energy transferred to the mechanical system per cycle

$$=\frac{Bp^2f^2A}{2\pi\ k_d^2}\int_{-k_d/2f}^{k_d/2f}\left\{t+\frac{k_d}{2f}\right\}^2\text{lg}_c\,\omega\sin\omega t\,dt$$

$$=\frac{Bp^2A\ \text{lg}_c}{2\pi}\left\{\frac{\sin\ [\pi k_d]}{\pi k_d}-\cos\ [\pi k_d]\right\}$$

and this will be a maximum when

$$(\pi^2k_d^2-1)\sin[\pi k_d]+\pi k_d\cos[\pi k_d]=0$$

or when k_d is 0.88, i.e., the value of contact closure period necessary for resonance.

In the case of the resonant scheme, the energy transferred to the mechanical system.

$$=\frac{2Bp^2A\omega\,\text{lg}\,f^2}{\pi^2k_d^2}\int_0^{k_d/2f}t^2\sin\omega t\,dt+$$

$$\frac{Bp^2A\omega\,\text{lg}}{2\pi(1-k_d)^2}\int_{k_d/2f}^{1/2f}(1-2ft)^2\sin\omega t\,dt$$

$$=\frac{Bp^2A\ \text{lg}}{\pi^2}\Bigg[\left\{\frac{1}{k_d}+\frac{1}{(1-k_d)}\right\}\sin\ [\pi k_d]+$$

$$\left\{\frac{1}{\pi k_d^2}-\frac{1}{\pi(1-k_d)^2}\right\}\cos\ [\pi k_d]-$$

$$\frac{1}{\pi k_d^2}-\frac{1}{\pi(1-k_d)^2}\Bigg]$$

and when k_d is 1/2, this becomes

$$0.147\ Bp^2A\ \text{lg}$$

The energy transferred to the magnetic field

$$=\frac{2Bp^2f^2At^2}{\pi k_d^2}\{\text{lg}+\text{lg}\cos\omega t\}$$

and when k_d is 1/2 this is a maximum when $t=1/4f$, so that maximum field energy

$$=0.1595\ Bp^2\ A\ \text{lg}$$

and

$$\frac{\text{energy transferred to the mechanical system}}{\text{energy transferred to the magnetic field}}=92.3\%$$

In the case of the conventional system, the energy transferred to the magnetic field is given by

$$\frac{Bp^2\ A\ \text{lg}_cf^2}{2\pi\ k_d^2}\left\{t+\frac{k_d}{2f}\right\}^2\{1+\cos\ [\omega t]\}$$

and when k_d is 1/2 this is a maximum when it is approximately $0.2/f$ so that the maximum field energy

$$=0.1695\ Bp^2\ A\ \text{lg}_c$$

The energy transferred to the mechanical system when k_d is 1/2 equals

$$0.1015\ Bp^2A\ \text{lg}_c$$

so that

$$\frac{\text{energy transferred to the mechanical (conventional) system}}{\text{energy transferred to the field of the conventional system}}=59.8\%$$

The relationship between the energy transferred to the mechanical (conventional) system for all values of contact closure periods and the maximum energy transferred to the magnetic field of the conventional system is shown in Fig. 7.

Considering only the energy provided by the actuating coil in the resonant system when k_d is 1/2 the energy transferred to the magnetic field is given by

$$\frac{2\ Bp^2Af^2t^2}{\pi}\{\text{lg}+\text{lg}\cos\omega t\}$$

the maximum value of which

$$=0.04\ Bp^2\ A\ \text{lg}$$

so that

$$\frac{\text{energy of the field of the resonant system}}{\text{energy of the field of the conventional system}}=\frac{0.04\ \text{lg}}{0.1695\ \text{lg}_c}$$

The value of lg/lg_c can be evaluated from the coefficients of the fundamental harmonic of force, viz.,

$$\frac{\text{lg}}{\text{lg}_c}=\frac{1.46}{\sqrt{1+\left\{\frac{\pi}{2}-\frac{2}{\pi}\right\}^2}}=1.1$$

and since the energy of the field ∝ {current}²

$$\frac{\text{current in the resonant system}}{\text{current in the conventional system}}=\sqrt{\frac{0.04(1.1)}{0.1695}}=0.51$$

The flux may also be resolved into harmonic components in order to evaluate the iron losses of the systems.

From Fig. 6(A)

$$\text{Flux}=\frac{Bp}{\pi k_d}\{t+\pi k_d\}$$

the sine coefficient being

$$\frac{Bp}{\pi^2k_d}\int_{-\pi k_d}^{\pi k_d}(t+\pi k_d)\sin nt\,dt$$

from which the fundamental

$$=\frac{2Bp}{\pi^2k_d}\{\sin\ [\pi k_d]-\pi k_d\cos\ [\pi k_d]\}$$

Similarly the cosine coefficient

$$=\frac{Bp}{\pi^2k_d}\int_{-\pi k_d}^{\pi k_d}(t+\pi k_d)\cos nt\,dt$$

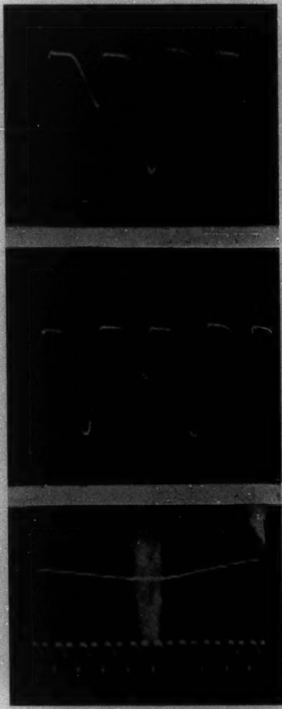


Fig. 8. Oscillograms of voltage waveforms (abnormal conditions)

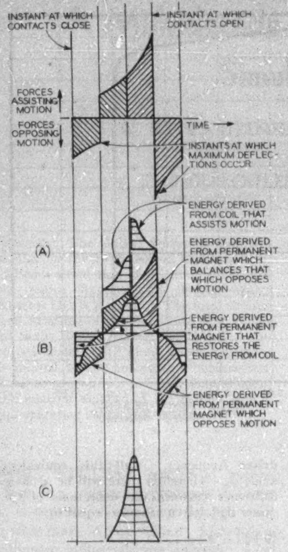


Fig. 9. Flux-time diagrams of

A—Permanent magnet
B—Combined coil and permanent magnet
C—Resultant

the fundamental being

$$\frac{Bp}{\pi^2 k_d}\{2\ \pi k_d \sin\ [\pi k_d]\}$$

When k_d is 1/2 one obtains for the peak harmonic flux a value of

$$\sqrt{\left(\frac{4Bp}{\pi^2}\right)^2+\left(\frac{2Bp}{\pi}\right)^2}=0.755\ Bp$$

In the case of the resonant scheme the flux is given by

$$\frac{Bpt}{\pi k_d}$$

so that the sine coefficient

$$=\frac{Bp}{\pi^2 k_d}\int_{-\pi k_d}^{\pi k_d} t \sin nt\ dt$$

the fundamental being

$$=\frac{Bp}{\pi^2 k_d}\{2 \sin\ [\pi k_d]-2\ \pi k_d \cos\ [\pi k_d]\}$$

When k_d is 1/2 this becomes

$$\frac{4Bp}{\pi^2}=0.406\ Bp$$

Considering the case when Bp is 2 kilogauss we have a peak value of flux of 1,510 gauss in the case of the conventional system and a peak value of 812 gauss in the case of the resonant scheme. The watts loss per pound at 500 cps (cycles per second) is found (in manufacturer's data for a particular grade of iron) to be 0.5 in the case of the conventional system and 0.18 in the case of the resonant scheme. At a frequency of 200 cps the values are found to be 0.14 and 0.06 respectively.

Additional losses occur in the conventional system because of the surge of current flowing through the coil resistance that is necessary to charge the shunt capacitor to the voltage of the supply, when the contacts close; see Fig. 5. With the tuning of the voltage waveform in the proposed resonant scheme this does not occur since the oscillatory voltage attains the same voltage as that of the supply when the contacts close.

Conclusions

It will be evident that the tuning of the voltage waveform as proposed without the use of a permanent magnet will have unsatisfactory results. The effect can be examined as in the following.

Consider the half-cycle: $-kd/2f$ to $kd/2f$; the flux is given by

$$\frac{2Bpf}{k_d}\,t$$

and the force

$$=\frac{4Bp^2f^2t^2}{k_d^2}\,\frac{A}{8\pi}$$

The mechanical energy derived during the half-cycle will be

$$=\frac{4Bp^2f^2 \lg\omega A}{k_d^2\pi}\int_{-k_d/2f}^{k_d/2f} t^2 \sin \omega t\ dt$$

$$=\frac{Bp^2f^2\omega \lg A}{2k_d^2\pi}\left[-\frac{t^2 \cos \omega t}{\omega}+\frac{2\,t \sin \omega t}{\omega^2}+\frac{2 \cos \omega t}{\omega^3}\right]_{-k_d/2}^{k_d/2}$$

and this will be found to be zero whatever vaue of contact closure period is employed.

It is therefore observed that only a small degree of asymmetry of the flux wave will enable oscillation to be obtained but the system will be very sensitive resulting in instability. This is indicated by the abnormal voltage waveforms shown in the oscillograms Fig. 8.

The use of a permanent magnet and tuned waveform as described effects a phase shift of the drive force, so that compared with the conventional system greater amplitudes of oscillation may be obtained (in some cases it will be doubled) for the same power input, or alternatively for the same amplitude the power input may be reduced (in some cases halved). Furthermore amplitudes exceeding those obtained at resonance can be achieved for all practical values of contact closure period whereas this can never be achieved in the conventional system.

The strength of the permanent magnet need have no critical value only in so far as that necessary to conserve all the counter-motion components of force. All the energy derived from the coil will be utilized in producing motion at this value, beyond which, very little further energy is derived, the energy expended by the permanent magnet in assisting the motion is counterbalanced by approximately the same amount opposing the motion. This is illustrated in Figs. 9(A) and 9(B), which serve to show how the coil flux is reimbursed. The maximum energy derived is shown in Fig. 9(C). It will therefore be observed that the strength of the permanent magnet can be increased further without upsetting the operation, and so realize greater starting forces. In some cases it may be preferable to increase the total coil input by the adoption of two coils connected in series, still using permanent magnets and tuned waveform. Two such arrangements are shown in Figs. 10 and 11. The coils are wound differentially so that when one is attracting the armature the other is restrained by the opposing permanent magnet flux. This double acting (pull-pull) arrangement also has the advantage that any external force (such as high values of

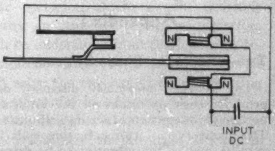


Fig. 10. Double acting drive system (Cantilever)

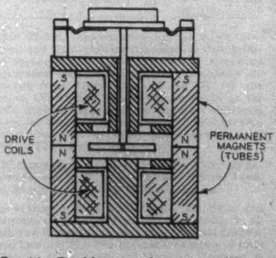


Fig. 11. Double acting drive system (Plunger)

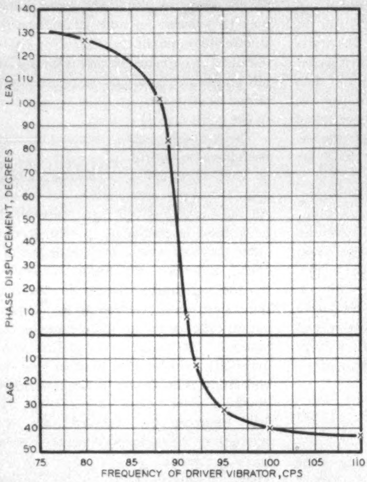


Fig. 12 (left). Phase displacement between the deflections obtained when one vibrator is used to drive another. Determined by calculation

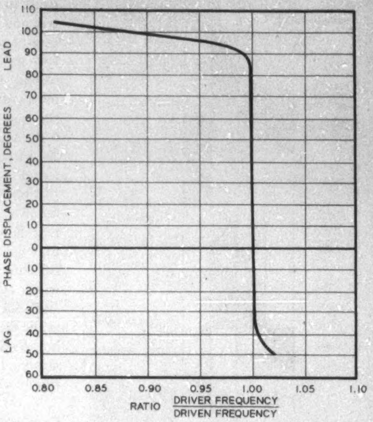


Fig. 13 (right). Phase displacement between the deflections obtained when one vibrator is used to drive another. Determined experimentally

acceleration) has a symmetrical disturbing influence on the system. Large starting forces will be required under such circumstances.

By tuning the waveform as described the oscillatory voltage spans the whole of the open period of the contacts so that the reversed voltage peak will be smaller than that obtained in the conventional system, thus insulating and radio interference problems are simplified.

Appendix I

Two- and Three-Phase Systems

The systems[1] previously discussed depend for their operation on the interconnection of the contacts of separate vibrators. The connection of two actuating coils in parallel enables the vibrators to be driven by the drive contact of one of them, but it gives rise to a two-phase oscillation and an explanation of this may be obtained by further consideration of the equation of motion y

$$=\frac{\sqrt{H_s^2+H_c^2}}{\sqrt{(S-\omega^2M)^2+r^2w^2}}\cos\,[wt-\theta-\psi]$$

where

$$\psi=\tan^{-1}\frac{r}{\frac{S}{\omega}-\omega M}$$

and

$$\theta=\tan^{-1}\frac{H_s}{H_c}$$

It is seen from this that if the mechanical characteristics of the two oscillating members are not identical, the frequency of oscillation will be governed by the characteristics of the driver vibrator and that since θ is the same for both vibrators when they are operated by their own drive contacts with equal contact closure periods the equivalent angle ψ for the driven vibrator must vary from that of the driver to satisfy the condition of operation at the driver frequency. Call this equivalent angle ψ_1. Therefore there will be a phase difference between the deflections of the driver and driven vibrators equal to

$$\{\theta+\psi\}-\{\theta+\psi_1\}$$

$$=180-\{\theta+\psi_1\}$$

A complete range of values can be evaluated as has been undertaken in Fig. 12 for a ratio of contact closure period of 0.5. Fig. 13 illustrates a similar relationship but determined experimentally. Thus, it will be observed that the graph has a very steep portion indicating that only a small difference in the mechanical characteristics of the driver and driven oscillating members is required to cause the phase displacement to be 90-degree lead. Furthermore complete transition from lagging to leading phase displacement can be achieved with small change in operating frequency.

References

1. A HIGH POWER VIBRATOR CONVERTOR, **R. Stewart Nunn.** *Electrical Energy*, London, England, vol. 2, no. 8, Aug. 1958, pp. 294–99.

2. APPLIED ELASTICITY (book), **S. Timoschenko, J. M. Lessells.** Westinghouse Technical Press, Pittsburgh, Pa., first edition, 1925 pp. 316–17.

A Multidecade Linear Reactor Simulator

J. W. SCHWARTZENBERG
NONMEMBER AIEE

J. M. FINAN
ASSOCIATE MEMBER AIEE

V. S. UNDERKOFFLER
NONMEMBER AIEE

R. N. BREY, JR.
NONMEMBER AIEE

Synopsis: The United States Government exhibit for the 1958 Geneva Conference included a simulation of the main portion of the entire Experimental Boiling Water Reactor (EBWR) plant. The total simulation consisted of two operating panels similar to the EBWR control panels and an animated model of the plant. This discussion is limited to the computer technique utilized to accomplish a linear simulation of the neutron kinetics of the reactor over a full operating range of 8 decades.

ol. 79
t. 1
960

THE NUCLEAR reactor is unique among industrial processes because it is capable of operating through such a wide range of flux values. Although a common industrial-type process might be capable of operating only through a range of 100 to 1, a nuclear reactor can easily operate through a range of 100,000,000 to 1, or 8 decades. This fact is responsible for the great difficulty that exists in adequately simulating the reactor throughout its entire operating range by means of analog computing equipment.

There are several possible methods available for solving the neutron kinetic equations of the reactor. For many applications, adequate results may be obtained by solving the linear form of the kinetic equations with a simple linear simulation.[1-7] This method of solution is limited to a maximum range of $2^1/_2$ to 3 decades by the characteristics of the analog equipment involved. Although this operating range is entirely adequate for many problems, the operation of EBWR covers a range of 8 decades which could not be performed by a simple linear simulation.

One possible method of solving the kinetic equations involves dividing the equations by the factor $1/n$, the inverse of the neutron flux signal, and solving for $(1/n)\,dn/dt$, the inverse period. This result is then integrated to obtain the natural logarithm of the output.[8] This method, which can successfully cover a range of 6 to 8 decades, is ideal for conducting short transient studies, such as start-up accidents, where the computer is run only for a short period of time and where a linear output signal is not required.

Several basic limitations, however, are present in this method. First, it is often necessary to obtain a linear flux signal. This becomes extremely difficult if a range of 8 decades is to be covered. Second, the analog computer circuit requires the use of an open-loop integrator. This is undesirable for a real time training simulator that may be run for many hours. A further limitation results because the computer circuit requires that a multichannel multiplier be placed in a high gain amplifier loop. When relatively short values of neutron mean lifetime are used, the loop gain becomes great enough to make a servomultiplier loop somewhat unstable. An electronic-type multiplier would perform satisfactorily but the circuit would require a 3-channel multiplier for even a 2-group approximation.

One additional method makes use of two linear simulators, each covering specified portions of the reactor output range. While one of the simulators is in the operating mode, the other is in the reset mode where it is constantly receiving its initial conditions from the operating simulator. When the first simulator reaches the end of its operating range, a switch-over is made and the two simulators interchange their modes of operation. The second simulator assumes the computing mode while the first simulator goes into the reset condition.

This switching may be repeated as often as necessary to permit the two simulators to cover the full range of the nuclear reactor output which may be 8 decades or more. The principle limitation of this method is the rather complex logic circuit and relay system required to perform the switching operation. However, for the simulation of the EBWR power plant, it was necessary to provide a linear output signal throughout the entire range of 8 decades. The problems associated with providing a linear signal from the logarithmic simulator, made this 2-channel simulator method the most desirable. Before considering the details of the 2-channel simulator, it will be helpful to describe briefly the operation of a simple linear simulator.

Linear Simulator

The dynamic response of many nuclear reactors can be described by the following set of differential equations which relate neutron flux and total reactivity.

$$\frac{dn}{dt}=\frac{\delta K n}{\ell^*}-\frac{\beta n}{\ell^*}+\sum_{i=1}^{n}\lambda_i C_i+S \tag{1}$$

$$\frac{dC_i}{dt}=-\lambda_i C_i+\frac{\beta_i n}{\ell^*} \tag{2}$$

where

n = thermal neutron density, neutrons/cm (per centimeter)^3
δK = reactivity $=(k_{eff}-1)/k_{eff}$
ℓ^* = mean effective lifetime of neutron in finite reactor, seconds
λ_i = decay constant, sec^{-1}
C_i = delayed neutron concentration of ith group
β_i = fraction of prompt neutrons delayed by ith group
$\beta=\Sigma\beta_i$
S = neutron source

Since the derivation and interpretation of these familiar equations has been covered in great detail elsewhere, no attempt will be made to explain them in this paper. It is sufficient for us to be able to construct an analog computer circuit that will properly obey these nuclear kinetic equations. This is done in Fig. 1 which shows a computer circuit for solving a 2-delay group approximation. The operation of the circuit, which is drawn in standard computer notation, can be understood from the following description. The two integrators on the left solve the two equations representing the effect of delayed neutrons.

Assuming that a signal proportional to $-n$ is available, the output of each of the integrators will be $(\ell^*\lambda_1/\beta_1)\,C_1$ and $(\ell^*\lambda_2/\beta_2)C_2$ respectively. The appropriate values of λ for the two groups are indicated by the coefficient potentiometer on the

PA

Paper 60-516, recommended by the AIEE Nucleonic and Radiation Instrumentation Committee and approved by the AIEE Technical Operations Department for publication. It is essentially abstracted from "Simulation of EBWR for the Geneva Conference," which was presented at the AIEE Nuclear Engineering and Science Conference, Cleveland, Ohio, April 6-9, 1959. Manuscript submitted December 4, 1958; made available for printing February 10, 1960.

J. W. SCHWARTZENBERG, J. M. FINAN, V. S. UNDERKOFFLER, and R. N. BREY, JR., are with the Leeds and Northrup Company, Philadelphia, Pa.

The authors wish to acknowledge the advice and assistance of J. M. Harrer, W. C. Lipinski, V. C. Hall, and E. Wimunc of Argonne National Laboratory, Lemont, Ill. The assistance of many of their colleagues at the Leeds and Northrup Company is also gratefully acknowledged.

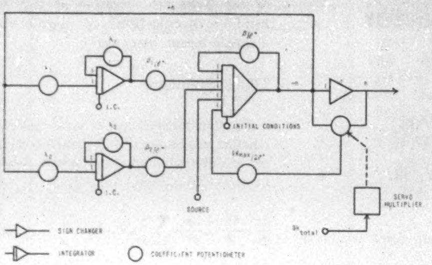

-n
λ1
λ1
β1/ℓ*
β/ℓ*
-n
n
I.C.
INITIAL CONDITIONS
λ2
λ2
β2/ℓ*
δkmax/2ℓ*
I.C.
SOURCE
SERVO MULTIPLIER
δktotal
SIGN CHANGER
INTEGRATOR
COEFFICIENT POTENTIOMETER

Fig. 1. Analog computer circuit to solve neutron kinetic equations

diagram. These delayed neutron signals are modified by the appropriate gain factors β_1/ℓ^* and β_2/ℓ^* to give signals proportional to $\lambda_1 C_1$ and $\lambda_2 C_2$. These signals, together with signals proportional to $-\beta n/\ell^*\ \delta K n$, and a source term S are then fed to an integrator which solves equation 1 and gives an output signal $-n$. Since this was the initial assumption, the circuit will correctly solve the desired equations. The sign change at the left of the circuit is necessary to provide both positive and negative signals to the servo-multiplier so that the product of δK and n may be obtained for both positive and negative values of reactivity.

The computer circuit shown in Fig. 1 requires an operational amplifier for each of the two delay groups shown because the appropriate delay equations are solved actively rather than by the more common passive method. This has been done so that when this circuit is extended to a dual channel simulator, the appropriate initial condition may be transferred between computers. If the passive method were used, it would be extremely difficult to charge one lag circuit from the corresponding one in the second computer.

Since the active method of simulation requires the use of an amplifier for each group of delayed neutrons, it was decided to use a 2-group approximation rather than the six groups that are actually present.

For the EBWR simulation the following constants were used:

$$\begin{array}{ll} \lambda_1 = 0.275\ \text{sec}^{-1} & \beta_1 = 0.0053 \\ \lambda_2 = 0.0263\ \text{sec}^{-1} & \underline{\beta_2 = 0.0020} \\ \ell^* = 6\times 10^{-6} & \beta = 0.0073 \end{array}$$

It was felt that this 2-group approximation gives a sufficiently accurate result to justify the reduction in computing equipment made possible by its use.

Description of Multidecade Simulator

The multidecade simulator incorporates two complete single-channel linear computers such as the one described in Fig. 1. One of the two computers is always in the operating or computing mode, while the other is in the reset mode, receiving its initial conditions from the operating computer. The technique by which this is accomplished is shown in Fig. 2. Here only the final integrator and sign changer of each computer is shown but similar circuitry is used for the two delayed neutron integrators.

Each integrator has two relays asso-

Vol.
Pt. 1
1960

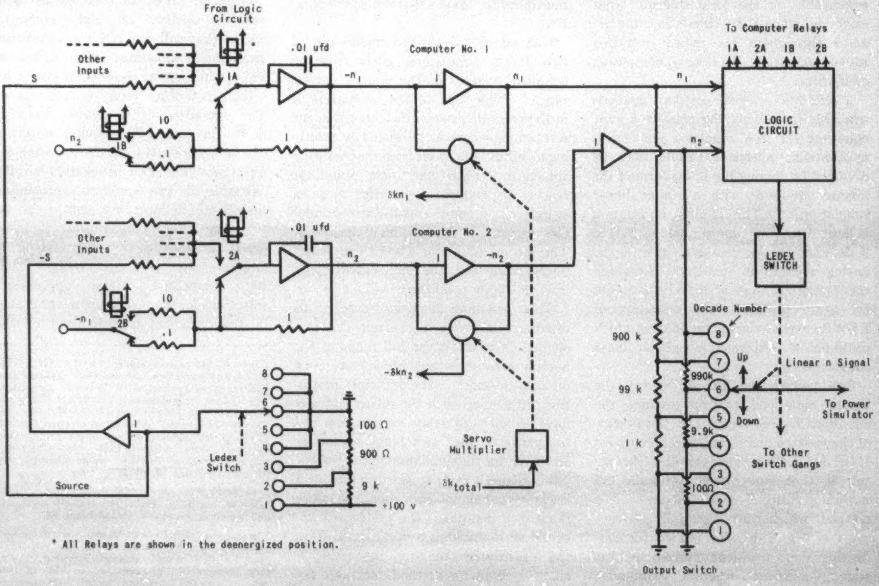

From Logic Circuit
Other Inputs
S
1A
.01 ufd
-n1
Computer No. 1
1
n1
n2
1B
10
.1
1
δkn1
Other Inputs
-S
2A
.01 ufd
n2
Computer No. 2
1
-n2
-n1
2B
10
.1
1
-δkn2
8
7
6
5
4
3
2
1
100 Ω
900 Ω
9 k
+100 v
Ledex Switch
Source
1
Servo Multiplier
δktotal
To Computer Relays
1A
2A
1B
2B
n1
n2
LOGIC CIRCUIT
1
LEDEX SWITCH
Decade Number
900 k
99 k
1 k
990k
9.9k
100Ω
Up
Down
Linear n Signal
To Power Simulator
To Other Switch Gangs
Output Switch
* All Relays are shown in the deenergized position.

Fig. 2. Simplified circuit of multidecade reactor simulator

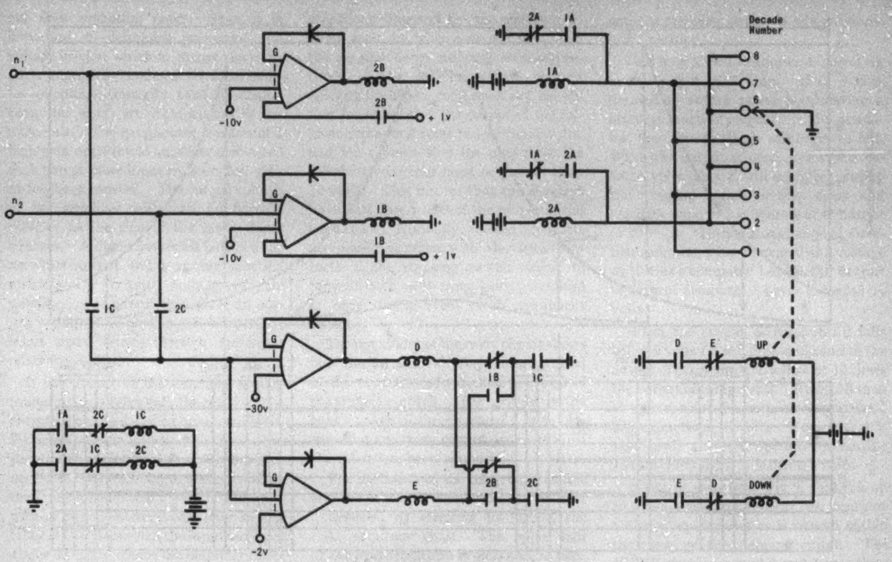


Fig. 3. Relay logic circuit

1.79
1
60

ciated with it, the *A* and the *B* relay. When the *A* relay is energized all input resistors are connected to the integrator and the circuit computes in the usual manner. When the *A* relay is de-energized, however, the summing junction of the integrator is transferred to a resistance network consisting of two input resistors, selected by the *B* relay, and a feedback resistor. A signal applied to the input lead of this network charges the feedback capacitor to a voltage proportional to the input signal and the ratio of the feedback to the input resistance. There are time lags associated with this charging process of 0.01 and 1.0 second for the final and delayed neutron integrators respectively which have essentially a negligible effect upon the performance of the computers.

Referring to Fig. 2, it can be seen that when relay *1A* is de-energized, computer 1 receives its initial conditions from computer 2. Likewise, when relay *2A* is de-energized, computer 2 receives its initial conditions from computer 1. Close inspection of the figure will reveal that the computer signals are arranged to be of opposite polarity. This is necessary because a signal reversal occurs in each computer amplifier. Therefore, $-n_1$ must be used as the initial condition signal for n_2 and, in a corresponding manner, n_2 must be used as the initial condition signal for $-n_1$.

In order that the two computers will be capable of simulating both increasing and decreasing flux signals, it is necessary for the computer that is in the reset mode to follow at an output voltage of either 1/10, or 10 times the output of the operating computer. When one computer reaches the top of its operating range, the other computer must be ready to operate with an output voltage that is lower by a factor of 10. In a similar fashion, when one computer reaches the bottom of its operating range, the other computer must be ready to operate with an output voltage that is higher by a factor of 10. These two conditions are provided by the action of the *B* relays.

When the *B* relay is energized, the ratio of the resistance network is 1/10, and the reset computer follows at a factor of 1/10 below the operating computer. When the *B* relay is de-energized, the ratio of resistance is 1/0.1, and the reset computer follows at a factor of 10 above the operating computer.

As a result of selecting these ratios, each of the computers is restricted to operate over a range of 1 decade. The remainder of the useful operating range of 3 decades is required to insure that the proper initial conditions are produced. If a choice had been made to restrict the simulation to cases where the power was only increasing, each computer could operate satisfactorily through a range of 2 decades. However, it was felt to be extremely important to be able to reduce as well as increase power level.

We have seen how the two linear computers are arranged to alternately compute and follow throughout the full range of the simulation. Since each computer will operate for 1 decade it is convenient to assign computer 1 to the odd-numbered decades and computer 2 to the even-numbered decades. It is then necessary to provide a means of control so that a usable output signal may be obtained from the two computers. First, an additional sign changer is required to provide neutron signals of the same polarity from both of the computers. This additional sign changer is placed in the output of computer 2 to give the output $+n_2$. When output signals of the same polarity are obtained they are then fed to the logic circuit and the output switching circuit.

The logic circuit determines the operation of the two computers. It has as its outputs the *A* and *B* relays that control the reset and compute modes and a bidirectional stepping switch which selects the decade that is being simulated.

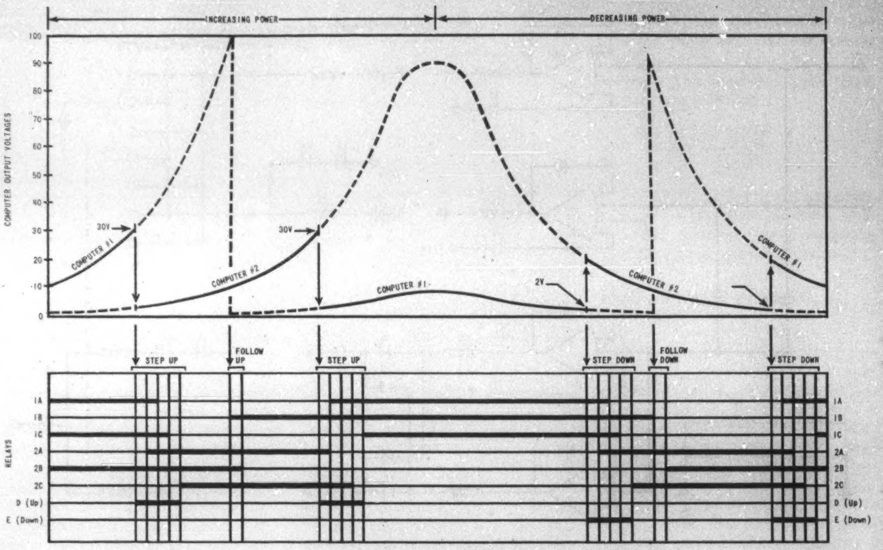


Fig. 4. Relay sequence chart for logic circuit

Vol.
Pt. 1
1960

When the limit of 1 decade is reached the switch will be stepped up or down to reach the desired decade. A complete description of the logic circuit will be given in the next section.

The output signals are also supplied to the output switching circuit. Each of the various decades of the simulators operating range are represented by the same voltage range in the computers. In order to obtain a full 8-decade linear output signal, it is necessary to switch the output signal lead alternately to the computer that is operating. It is also necessary to insert the proper scale factor for the decade involved. When operating in decade 8 for example, the output is directly connected to computer 2. In decade 7, however, the output lead is connected to computer 1 through a gain factor of 1/10. Similarly in decade 6, the output is connected to computer 2 through a gain factor of 1/100. This output switching process converts this combined analog-digital output signal of the two computers into a continuous analog signal. Only the output switch for the power simulation is shown in Fig. 2; however, similar switching arrangements are used for the outputs to the linear N and log N recording channels.

A further study of Fig. 2 will show that the magnitude of the source term S is modified by the position of the stepping switch. This is necessary, since each decade of the simulation represents many different neutron levels by the same voltage ranges. Therefore, it is necessary to represent a constant neutron source term by a varying voltage. The strength of the source is reduced by a factor of 10 for each decade the simulator is operating above the source signal. Since the two computers produce signals of opposite polarity, a sign changer is required to produce source signals of opposite polarities. The source level is set equal to zero after the first 3 decades of operation.

In this section we have described how the two individual computers are operated together to produce a common output signal over a range of 8 decades as a function of the input reactivity signal δK. The next section will describe how the logic circuit selects the proper decade of operation.

Logic Circuit

The basic operation of the two computers that make up the multidecade simulator is controlled by the logic circuit shown in Fig. 3. As previously stated, in order that the simulator can be capable of both increasing or decreasing power level, the ranges during which each computer operates is restricted to 1 decade. A convenient voltage selection for the operating range is from 2 volts to 30 volts. The remainder of the useful computer output range is reserved for the reset mode where the proper initial conditions are inserted.

When a computer in the operating mode reaches a level of 30 volts, the stepping switch is advanced one position causing the alternate computer to begin operating at a level of 3 volts. This process can be repeated over and over again. When an operating computer reaches a level of 2 volts, the stepping switch is stepped down one position causing the alternate computer to begin operating at a level of 20 volts. This process may also be repeated throughout the range of the simulation.

The condition of the computer in the reset mode is controlled from the voltage level of the operating computer. When the voltage of the operating computer exceeds 10 volts, the reset computer is prepared to follow an up step and therefore is following a voltage that is 1/10 the output of the operating computer. When the voltage of the operating computer becomes less than 9 volts, the reset computer is then prepared to follow a down step and therefore follows at a voltage that is 10 times the output of the operating computer. If these basic rules of operation are obeyed by the logic circuit, the two computers will function correctly throughout their operating range. Since it is important to determine accurately the voltage level at which to change the computer's operating modes, several comparator circuits, commonly used in analog computer work, are incorporated in the logic unit. The comparator consists of a high gain operational amplifier connected with two or more input resistors but with no feedback resistor. The output voltage of the amplifier will then flip-flop its polarity as the sign of the input signal changes. A diode connected between the amplifier output and summing junction restricts the output voltage to one polarity. The circuit then gives an output voltage of either 0 or 100 volts as the input signal passes through the input reference voltage.

If the output of the amplifier is then connected to a relay coil, the relay will be energized by the specified input condition. For example, the lowest of the four comparators shown at the left of Fig. 3 is connected so that the relay *E* energizes when the input drops below 2 volts. The comparator directly above is connected so that relay *D* energizes when the input increases above 30 volts. Since the amplifiers have a very high gain, the exact point of switching is determined by the input resistors and input voltages and is relatively independent of relay characteristics.

The relay circuit shown in Fig. 3 performs just the operations that were stated at the beginning of this section. When n_1 is greater than 10 volts, relay *2B* is energized by the comparator circuit. The additional 1-volt signal carried over a normally open contact of *2B* prevents the relay from dropping out until n_1 is less than +9 volts. The second comparator circuit performs a similar function on relay *1B* as input n_2 changes value.

The input signal to the lower two comparators is broken over normally open contacts of *1C* and *2C*. These relays are arranged to repeat the position of the operate-reset relays, *1A* and *2A*. Their circuits, however, include a normally closed contact which prevents both *C* relays from being energized when both *1A* and *2A* are energized. This is important since, at the time of a switch-over, it is necessary to have both computers in the operate mode and correspondingly both *1A* and *2A* energized, for a brief instant to insure a continuous output signal.

By means of the *C* relay network, the lower two comparators are connected to the computer in the operate mode. The third comparator circuit is energized when the input exceeds 30 volts, and the fourth comparator circuit is energized when the input is less than 2 volts. If the conditions imposed by the network in series with the relay coils are met, then the up and down stepping coils respectively of the stepping switch can be energized. These conditions are simply that for an up step the output of the following computer must be less than 9 volts, and for a down step the output of the following computer must be greater than 10 volts. This insures that the *B* relays have had time to insert the proper initial conditions. Normally, closed contacts are placed in series with the open contacts in the stepping switch circuit to prevent both coils from being energized at once in the event of an equipment failure.

The remainder of the relay circuit shows how the *2A* and *1A* relays are energized at the even and odd decade positions of the stepping switch. The contacts in the holding path of the *1A* and *2A* relays insure that one relay cannot drop out until the other has been energized.

The operation of the relay logic circuit and the complete simulation can be best understood by referring to Fig. 4, the relay sequence chart. The upper half of the figure indicates typical analog voltages that might be present in the computers while the lower half shows the condition of the various relays corresponding to the analog voltages. A heavy line indicates that the relay is energized.

Starting at the left of the figure, assume that computer 1 is operating at a level of 10 volts and that computer 2 is following at a level of 1 volt. At this point relays *1A*, *1C*, and *2B* are energized; relays *1A* and *1C* energized indicating that computer 1 is operating, and *2B* energized indicating that computer 1 has an output of 10 volts or more. As a positive reactivity input is supplied to the simulator, the output begins to increase in an exponential fashion. When the output of computer 1 reaches 30 volts, the comparator circuit fed through *1C* energizes the *D* relay and causes the stepping switch to advance position. This causes *2A* to energize, *1A* to drop out, *1C* to drop out, *D* to drop out, and *2C* to energize. Computer 2 is now operating and computer 1 following at a voltage that is 10 times the output of computer 2.

When the output of computer 2 exceeds 10 volts *1B* is energized, causing the output of computer 1 to follow at 1/10 the amount of computer 2. When computer 1 drops to 1 volt, relay *2B* drops out. The simulator is now ready to step up in a similar manner except this time computer 2 is connected to the comparator circuit through relay *2C*. When computer 2 reaches 30 volts the process repeats and the stepping switch is advanced another position.

This now leaves computer 1 operating at an output of 3 volts. If it is now decided to reduce power by inserting a negative reactivity signal into the simulator, the output voltage will begin to fall. When the output of computer 1 falls below 2 volts relay *E* will energize, causing the stepping switch to step down and placing computer 2 in operate at an output of 20 volts. During this change of direction computer 2 has remained at a voltage of 10 times computer 1 since the output voltage of computer 1 never exceeded 10 volts.

When the output of computer 2 falls below 9 volts *1B* will drop out, causing the output of computer 1 to follow at 10 times the output of computer 2. Relay *2B* then energizes, since the output of computer 1 is greater than 10 volts. Computer 1 is now ready to begin operating when the level of computer 2 falls below 2 volts.

It can be seen from Fig. 4 that this up or down switching process can continue for as many decades as is desired within the range of the stepping switch. The EBWR simulator used 8 switch positions corresponding to 8 decades of operation, but this could have been increased, if necessary. When the top of the range is reached, it is simply necessary to open the circuit to the up coil of the stepping switch to prevent further switching. The presence of the source term prevents the output of the simulator from decreasing below the lower end of the operating range.

Conclusions

The dynamic responses of the simulator closely approximate the actual performance of EBWR. Transients do occur in output during computer switching. The transients are larger when the switching is in the down direction than in the up direction because of the time constant associated with charging the delay neutron integrators to the correct initial conditions. However, these transients did not detract from the performance of the simulator since the characteristics of the two computers were closely matched to minimize any output offset.

Additional information on this subject can be found in the list of references.

References

1. ELECTRONIC PILE SIMULATOR, P. R. Bell, H. A. Straus. *The Review of Scientific Instruments*, New York, N. Y., vol. 21, no. 8 Aug. 1950, pp. 760–63.

2. A PORTABLE ELECTRONIC PILE KINETIC SIMULATOR, Walter Pagels. *AIEE Transactions* vol. 70, pt. II, 1951, pp. 1422–26.

3. Oak Ridge National Laboratory Reactor Controls Computer, **J. J. Stone, E. R. Mann.** *ORNL-1632*, Oak Ridge National Laboratory, Oak Ridge, Tenn., Apr. 20, 1954.

4. The HRE Simulator, **D. M. Collier, L. A. Meeks, J. P. Palmer.** *ORNL-1572*, Oak Ridge National Laboratory, Oak Ridge, Tenn., Sept. 24, 1954.

5. A Nuclear Reactor Simulator for Teaching Purposes, **L. Orr, W. Kerr, H. J. Gomberg.** *Electrical Engineering*, vol. 75, no. 4, Apr. 1956, pp. 364–67.

6. Control of Nuclear Reactors and Power Plants (book), **M. A. Schultz.** McGraw-Hill Book Company, Inc., New York, N. Y., 1955, pp. 282–302.

7. A Simulator for Nuclear Reactor Operator Training, **J. W. Schwartzenberg.** *ISA Journal*, Instrument Society of America, Pittsburgh, Pa., vol. 4, no. 9, Sept. 1957, pp. 369–73.

8. Nuclear Reactor Start-Up Simulation, **J. P. Franz, N. F. Simcic.** *Transactions*, Professional Group on Nuclear Science, Institute of Radio Engineers, New York, N. Y., vol. PGNS-4, no. 1, Mar. 1957, pp. 11–14.

9. Analysis of Experimental Power-Reactivity Feedback Transfer Functions for a Natural Circulation Boiling Water Reactor, **J. A. DeShong, W. C. Lipinski.** *ANL-5850*, Argonne National Laboratory, Lemont, Ill., July 1958.

A Dual-Channel Reactor Protection System for Nuclear Power Plants

ALLEN S. BARTU
ASSOCIATE MEMBER AIEE

Synopsis: A dual-channel reactor protection system similar to the one that was developed for the Dresden Nuclear Power Station is described. This protection system utilizes commercially available components in coincidence circuits to achieve the system reliability that is required for power plant service.

THE FUNCTION of the reactor protection system is to prevent the reactor from operating in a manner that is dangerous to itself or its environment, without jeopardizing its output at any time when it is operating safely.

In a nuclear power plant, one is faced with the marriage of two opposing philosophies. Traditionally, reactor protection has been designed to shut the reactor down if there is any doubt about plant safety. Power plant protection, however, has traditionally been designed to keep the plant running until it is definitely established that a dangerous condition exists. Both operating philosophies are well founded and the reactor protection system for a nuclear power plant must satisfy the basic requirements that led to both of these opposing philosophies.

During the design of the Dresden Nuclear Power Station, a dual-channel reactor protection system was developed. The system described herein is essentially a generalization of that design and is believed to be applicable to many types of nuclear power plants.

The reactor protection system which is discussed includes the sensing elements, logic functions, and actuator controls as shown in Fig. 1. It should be noted, however, that two other critical steps exist between the incidence of an unsafe condition and a shutdown. The unsafe condition must be sensed by the primary element, and the control rods must actually insert part of their negative reactivity before a scram has really taken place.

Basic System Characteristics

The important characteristics of a reactor protection system can be defined as speed of response, operational reliability, freedom from nuisance operations, and provisions for trouble monitoring, testing, repair, and maintenance. The relative importance of each of these characteristics is not the same, but all are important and a serious weakness in any one is a significant weakness in the protection as a whole.

Speed of Response

The speed of response for the whole scram system (including sensing of the unsafe condition and enough reduction in reactivity to put the reactor on a negative period) must be adequate to control safely some limiting potential accident. The limiting accident is usually the start-up accident,[1] which is determined by the nuclear characteristics of the reactor and the maximum reactivity insertion rate attainable from the control system.

The delay due to the protection system, as described herein, is only a portion, and often a small portion, of the total delay. The delay involved in getting the control rods to move far enough to produce significant negative reactivity is a major portion of the total delay, and the delay involved in sensing an unsafe condition by the primary sensing element may be a serious delay in some cases.

Operational Reliability

If the protection system has adequate speed of response, its prime characteristic is operational reliability. This can be defined as the ability to cause a reactor shutdown when it is required. Failure probabilities as low as one in a million are necessary for some situations.

Nuisance Operations

A nuisance operation can be defined as a protection-system operation which causes a shutdown of all or part of the plant when there is no danger. The most common example of nuisance operation is a nuisance scram.

Many of the reactor protection systems in use today cause a relatively large number of nuisance operations when compared to what is generally considered acceptable for a modern power plant.

Trouble Monitoring, Testing, and Maintenance Provisions

The ability to annunciate the failure of any of the components of the reactor protection system will greatly affect the operational reliability and the nuisance failure rate of the system. This is because the probability that the protection system will fail to operate when required is a function of the time that it takes to detect and repair a component failure. Somewhat the same thing applies to the nuisance failure rate but it is greatly affected by the particular protection system design.[2]

Testing can be considered a method of extending the trouble-monitoring capability by means of periodic checks. Ideally, the trouble monitor would check all components continuously and no routine testing would be required, but this is generally not practical. The testing pro-

---

Paper 60-78, recommended by the AIEE Nucleonic and Radiation Instrumentation Committee and approved by the AIEE Technical Operations Department for presentation at the Nuclear Engineering and Science Conference, Cleveland, Ohio, April 6–9, 1959. Manuscript submitted November 4, 1959; made available for printing November 23, 1959.

Allen S. Bartu is with the General Electric Company, San Jose, Calif.

All those who contributed to this development are too numerous to mention but special credit should be accorded I. M. Jacobs for developing some of the basic concepts, and K. F. Hempstead, of the Commonwealth Edison Company, for carrying the design through to completion.

Vol.
Pt. 1
1960

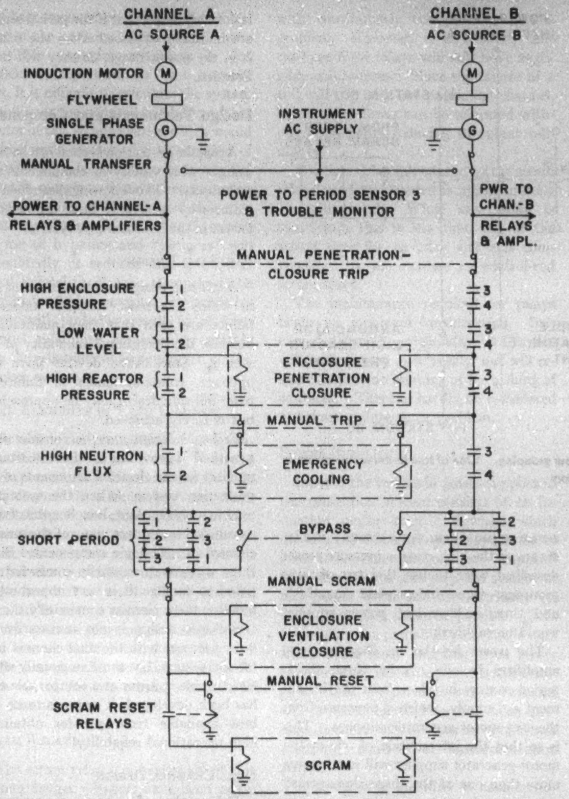

CHANNEL A
AC SOURCE A
INDUCTION MOTOR
M
FLYWHEEL
SINGLE PHASE GENERATOR
G
MANUAL TRANSFER
POWER TO CHANNEL-A RELAYS & AMPLIFIERS
CHANNEL B
AC SOURCE B
INSTRUMENT AC SUPPLY
POWER TO PERIOD SENSOR 3 & TROUBLE MONITOR
PWR TO CHAN.-B RELAYS & AMPL.
MANUAL PENETRATION-CLOSURE TRIP
HIGH ENCLOSURE PRESSURE
LOW WATER LEVEL
HIGH REACTOR PRESSURE
ENCLOSURE PENETRATION CLOSURE
MANUAL TRIP
HIGH NEUTRON FLUX
EMERGENCY COOLING
SHORT PERIOD
BYPASS
MANUAL SCRAM
ENCLOSURE VENTILATION CLOSURE
MANUAL RESET
SCRAM RESET RELAYS
SCRAM
1
2
3
4

Fig. 1. One-line diagram of a typical Dresden-type dual-channel reactor protection system

visions should give complete coverage of those areas that have been omitted by the trouble monitor, so that no practical, unsafe failure could be left undetected.

To keep the detection and repair time for a failed component to a minimum, provisions for quick and easy repair are very important. The ability to maintain and repair components with the reactor plant in operation is particularly important in a power plant where the reactor may be expected to operate for long periods without a shutdown.

Dual-Channel System

The Dresden-type dual-channel reactor protection system meets the reliability requirements by making use of coincidence techniques rather than by using extremely reliable components. The failure of any one component or power supply is assumed to be a normal design condition. The system is designed so that, in nearly all cases, such failures will not cause an operational failure or a nuisance operation, but will trip one protection channel with the other channel still intact.

Fig. 1 shows a 1-line diagram of a typical Dresden-type system as it might be applied to a boiling-water reactor. The sensed conditions will be different for different types of reactors, but the basic design should be applicable to a variety of reactor types.

The system consists essentially of two grounded 120-volt a-c circuits which are supplied from two induction-motor-driven synchronous generators with flywheels to provide a few seconds of stored energy.

Each numbered contact shown in Fig. 1 symbolizes a scram relay. A coil of each relay is supplied with power from an "independent" primary sensor, so each of these contacts may be thought of as symbolizing a primary sensor. The pairs of holding coils shown each symbolize one or more pairs of actual holding coils or solenoid valves. These coils or valves are mechanically interconnected in such a way that both must be de-energized before a protection system operation takes place.

Two-of-Four-Sensor Grouping

The high-reactor pressure sensor shown in Fig. 1 is typical of the two-of-four-sensor grouping which forms the basis for this dual-channel system. A typical scram relay circuit for one of these sensors is shown in Fig. 2

If the reactor pressure should reach the trip level, the pressure switches 1, 2, 3, and 4 will open their contacts. This will de-energize their respective scram relays and open the four high-reactor-pressure contacts shown in Fig. 1. These contacts will de-energize everything which is shown below them, including the holding coils controlling the ventilating ducts and the scram valves.

If any one of the switches or relays should fail to open its contacts, the other relay in its protection channel will open the circuit and the scram will still be accomplished.

If one switch or relay in each channel should fail to operate, the other two relays will still cause the required scram. This allows one switch in channel *A* and one in channel *B* to be connected to the same manifold without the possibility of a single plugged manifold's preventing a required scram. The fact that two manifolds may be used safely instead of three or four can result in significant savings in piping costs in a high-pressure system.

If any one of the high-reactor-pressure sensors should accidentally open its contacts, or have its circuits open-circuited or short-circuited, its associated scram relay would be de-energized. This would interrupt the current to one of each pair of ventilation-duct and emergency-cooling holding coils and one of each pair of scram solenoids, but the other holding coils and scram solenoids would remain energized and prevent a nuisance scram. The annunciator and trouble indicator light associated with this sensor would be energized so as to identify positively the circuit that caused the protection channel to trip. If the failed component were in an accessible location, it could be replaced or repaired without disturbing the plant output.

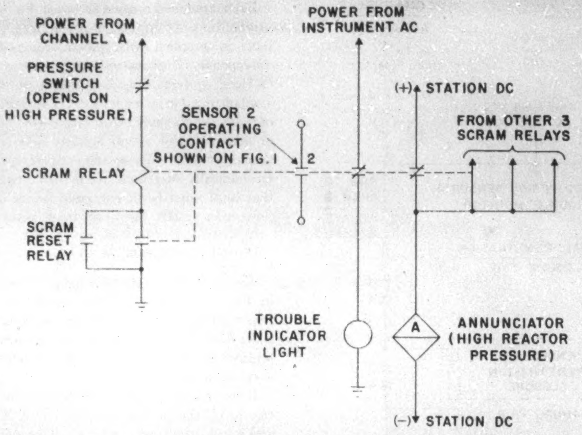

POWER FROM CHANNEL A
PRESSURE SWITCH (OPENS ON HIGH PRESSURE)
SCRAM RELAY
SCRAM RESET RELAY
SENSOR 2 OPERATING CONTACT SHOWN ON FIG. I
2
POWER FROM INSTRUMENT AC
TROUBLE INDICATOR LIGHT
(+) STATION DC
FROM OTHER 3 SCRAM RELAYS
A
ANNUNCIATOR (HIGH REACTOR PRESSURE)
(−) STATION DC

Fig. 2. Typical scram relay circuit for a two-of-four grouping. One of four high-reactor-pressure sensors

If the power output of one of the motor-generator sets were interrupted or short-circuited, the power to the ion chambers, amplifiers, relays, holding coils, and solenoids which are associated with that protection channel would be de-energized. Since no components in the other protection channel are supplied from this source, the other channel would still be operative and the reactor would continue to operate.

Two-of-Three-Sensor Grouping

The two-of-three grouping of sensors may be used with a dual-channel system when the saving of one sensor justifies the added complexity, cost, and/or the higher nuisance failure rate. The operational failure probability for the two-of-three grouping is nearly the same as for the two-of-four grouping.

The short-period sensor shown in Fig. 1 is typical of the two-of-three grouping. Fig. 3 shows a typical scram relay circuit for this sensor.

If reactor period reaches the trip level, the short-period sensors 1, 2, and 3 will de-energize their respective relays. The twelve short-period contacts shown in Fig. 1 will open and de-energize the holding coils controlling the ventilating ducts and the scram valves.

If any one of these sensors should fail, the other two will still open both circuits and cause the required scram. However, these three sensors must be completely independent because a failure which could affect any two might prevent a required scram or cause a nuisance scram. For instance, if the high-reactor-pressure sensor described were to use this two-of-three grouping, three independent manifolds and three independent power supplies would be required.

The power for the ion chambers and amplifiers in one of the three short-period sensors is taken from the instrument a-c supply, which is separate from the two motor-generator supplies. This is so that the interruption of one of the motor-generator supplies will not involve more than one of the three short-period sensors, and cause a nuisance scram. This short-period sensor will experience more nuisance operations than the other two because of voltage disturbances on the instrument supply However, this is not very important if the period sensors are normally bypassed when the reactor is in the power range, as they will be at Dresden.

Design Techniques and Components

A reactor protection system can be built using a wide variety of components and techniques. The following discussion explains why certain of these are selected for use in the Dresden-type system.

Fail-Safe Design

A fail-safe design can be defined as one in which the most common modes of failure are such that they maintain or increase the operational reliability of the system. Since some devices have two opposite, common modes of failure, a truly fail-safe design may be approached, but is rarely achieved.

Fail-safe design may be considered as a method of providing an effective trouble monitor for the electric components of the protection system. When the system is used in a power plant, long lengths of wire are needed to connect many of the sensing elements to the logic component. Since these wires are necessarily somewhat exposed to damage, it is very important to monitor these circuits continually.

Fail-safe design is not necessarily required for use with the dual-channel protection system, but until an equally effective trouble monitor at a comparable cost has been developed, it appears to be the best available technique for obtaining high operational reliability.

Dual-Channel Design

The number of nuisance operations is reduced by the use of a second, identical fail-safe channel which also must be de-energized in order to produce a scram.

The operational reliability appears to

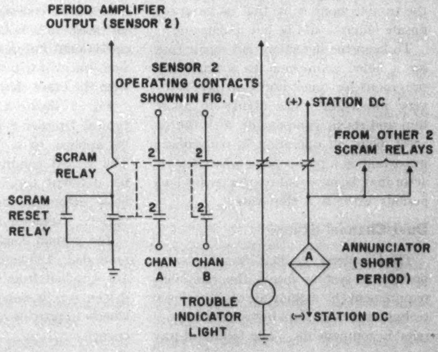

PERIOD AMPLIFIER OUTPUT (SENSOR 2)
SENSOR 2 OPERATING CONTACTS SHOWN IN FIG. I
SCRAM RELAY
SCRAM RESET RELAY
2
2
2
2
CHAN A
CHAN B
TROUBLE INDICATOR LIGHT
(+) STATION DC
FROM OTHER 2 SCRAM RELAYS
A
ANNUNCIATOR (SHORT PERIOD)
(−) STATION DC

Fig. 3. Typical scram relay circuit for a two-of-three grouping. One of three short-period sensors

have been reduced by a factor of two because of the addition of the second channel. This apparent reduction is not considered important for two reasons: First, it is difficult to calculate the system reliability or the required reliability within a factor of ten, so any system which would become unsatisfactory by a reduction of reliability by a factor of two is already too near the margin. Second, because accidental tripping of one channel due to a component failure is not in itself serious, the use of bypasses and "jumpers" will undoubtedly be reduced and the operational reliability of each channel will be increased. This somewhat intangible increase will probably result in a dual-channel reliability considerably greater than the reliability of an equivalent single-channel system.

DUAL ELEMENTS IN EACH CHANNEL

To reduce the probability of an operational failure due to an "unsafe" failure of a component, two elements are used in each channel in such a way that both must fail before the channel will be unable to scram from any one scram condition. These two elements are physically located so that the possibility of damage by a single accident is small. They are also installed in such a manner that a single failure, such as plugging of a single pressure line or control-rod shadowing of a single ion chamber, cannot affect more than one of the two sensing elements in one channel.

SCRAM RELAYS

The scram relay is a basic part of this system design. Relays have been justly criticized for their performance in some applications, but the manner in which a properly selected relay is likely to fail is quite well established.

The application of relays to a reactor protection system presents few significant problems and the selection of a relay which will perform satisfactorily should be quite simple.

Proper dust protection should easily reduce to a negligible value the probability of sticking. The probability of a contact welding is very small when a properly sized relay is supplying a resistive or inductive load. The expected number of operations is so small that fatigue problems are negligible.

The maximum currents involved are approximately 5 amperes a-c. This is well within the rating of the General Electric Company type *HFA* relays which were selected for use at Dresden.

Because of the thermal aging of the coils, the operation of the scram relays with continuously energized coils is a problem. However, it is believed that the type *HFA* relays will still have negligible coil failures. Since the failure of a coil will trip only one protection channel, a few coil failures can be tolerated without appreciably reducing the plant reliability.

The time delay introduced by the scram relays used at Dresden is approximately 10 milliseconds. While any delay is significant, this is less than 5% of the scram time for Dresden and was compensated by an increase in control-rod scram speed.

The maintenance required for relays is also an important consideration. It is expected that the scram relays at Dresden will be inspected once yearly, but will require little or no cleaning or burnishing of contacts. This can hardly be considered a serious maintenance problem.

POWER SUPPLIES

Each of the two main power supplies to the protection system consists of an induction motor with a flywheel which drives a 120-volt single-phase synchronous generator. The motors are supplied from the station auxiliary a-c system at two points which are as electrically remote from each other as possible. This is to reduce the probability of a single electrical disturbance affecting both power supplies when some a-c power is still available. If all a-c power is lost, the reactor will probably have to be shut down anyway because of the loss of important pumping, control, or instrumentation power.

A flywheel is included in each motor generator set to provide enough stored energy so that a short, auxiliary-power interruption will not scram the reactor. A practical amount of stored energy allows the protection system to continue to operate for 3 to 5 seconds after all power is interrupted.

Excitation for the synchronous generator is provided by a direct-connected exciter to provide maximum reliability.

The use of 120 volts a-c provides a good compromise between size of wire in relay and solenoid coils and current-interrupting and film-drop problems in switch and relay contacts. This is also a convenient voltage for supplying the ion-chamber and flux-amplifier power supplies. One side of each of the protection circuits is grounded so as to remove the possibility of two accidental grounds causing a sensor to be bypassed.

By means of a manual transfer switch, instrument a-c power can be used in place of the synchronous-generator output while one motor-generator set is being serviced or repaired. This temporary mode of operation will result in a somewhat higher nuisance failure rate. However, this condition can probably be tolerated for considerable periods so long as the other motor generator set is working properly. (The third short-period sensor will be transferred from the instrument a-c supply to the running motor generator set for this abnormal mode of operation.)

TROUBLE MONITORING, TESTING, AND MAINTENANCE PROVISIONS

Trouble monitoring for most of the electric system is provided by additional normally closed contacts on the scram relays as shown in Figs. 2 and 3. Since the protection circuits have one side grounded, either a short or open circuit will result in a dropout of the scram relay. This will cause tripping of one protection channel and will operate the trouble indicator light and an annunciator. The annunciator provides the operator with an audible and visual identification of the scram function involved, and the trouble indicator light positively identifies the circuit involved so that a maintenance mechanic can be sure he is working on the correct circuit.

Since the most common source of trouble will probably involve open or short circuits, this trouble monitor is expected to indicate and annunciate most protection system troubles.

Since the primary elements are transducers, they cannot be completely tested electrically. With a dual-channel system, however, routine operational checking of primary elements is possible without interrupting plant operation. For example, if the high-reactor-pressure censors are to be tested, one of the switches can be isolated from the reactor by closing the associated root valve, and an external pressure can be applied to the pressure switch until its protection channel is tripped. The trip setting and the operability of the switch and its associated circuits will both be checked by this test. After the test has been completed, the protection channel can be reset and the remaining switches tested one by one in a like manner.

All sensors which are in accessible locations and can be provided with an artificial primary signal can be operationally tested in a similar manner.

Testing of the scram mechanism will have to be accomplished by performing an actual scram. However, if it should be necessary, individual control rods or groups of rods can be scrammed by individually de-energizing their control valves with the remainder of the reactor in operation.

Since there are two independent, protection channels, either channel can be de-energized for maintenance without disturbing the other channel or the output of the plant.

The location of components and circuits at Dresden is such that there is little likelihood of a mechanic's accidentally disturbing the operating channel while testing or maintaining the other channel.

Conclusions

The Dresden-type dual-channel protection system appears to be able to meet the operational reliability required for reactor protection while still maintaining the nuisance failure rate low enough to be acceptable in a modern power plant. A design has been completed using proved, commercially available components in conventional circuits which lead to a high degree of predictability of performance. The proof of the design, however, still awaits the actual operating experience which will be obtained when Dresden Nuclear Power Station starts producing power.

References

1. PRINCIPLES OF NUCLEAR REACTOR ENGINEERING (book), S. Glasstone. D. Van Nostrand Company, Inc., Princeton, N. J., 1955, pp. 352–53.

2. SAFETY SYSTEM FOR NUCLEAR POWER REACTORS, I. M. Jacobs. *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 76, Nov. 1957, pp. 670–73.

Transients in Logarithmic Count-Rate and Period Meters

BRUCE B. BARROW
NONMEMBER AIEE

ROBERT MAITLAND
NONMEMBER AIEE

IT HAS BEEN found desirable and practical, in designing automatic control devices for starting nuclear reactors, to measure flux transients at extremely low flux levels, using the output from neutron-counting circuits. In general, a logarithmic count-rate circuit is used, and the logarithmic indication of flux thus obtained is differentiated electrically to yield an indication of the period of the reactor flux.

The special problem that arises when a period meter is built around a logarithmic count-rate meter is random fluctuation of the meter output, or *noise*. This noise is inseparably associated with the quantity being measured, for it arises in the randomness with which neutrons are detected in a fission counter or equivalent device. To reduce this noise requires, in general, measurement over a larger number of neutron detections, which means measurement over a longer time, i.e., slower transient response, assuming that counter sensitivity cannot be increased. It is possible, however, to achieve a good compromise between quick response and low noise, and the procedure for achieving such a compromise has been discussed in earlier papers.[1,2]

This paper cannot easily be studied apart from its predecessors. To write it otherwise would have required excessively long repetitions. This paper begins, therefore, where the others ended. Several classes of differentiating networks are analyzed in connection with the general problem of differentiating a signal while filtering out undesired wide-band noise. An indication of the relative quality of performance of these various networks is given, and confirmed quantitatively for the log-diode period meter. Experimental measurements of transient performance are reported, and these confirm the predictions from theory. The experiments involved both a simple log-diode circuit and an entire period meter. Transient inputs were obtained from an exponential-current generator and, later, from a reactor simulator operated with various reactivity inputs.

Differentiating-Network Design

GENERAL PROPERTIES

At least two types of logarithmic count-rate circuits have been used in period meters. The first of these, the log-diode circuit, consists essentially of a logarithmic diode, usually a vacuum diode operated in the logarithmic region of its current-voltage characteristic, in parallel with a capacitor. An electric charge is dumped onto this capacitor each time a neutron is detected, and the current leak through the diode causes a voltage that is more or less proportional to the logarithm of the counting rate. The important properties of this circuit are known.[2]

The second type of logarithmic count-rate circuit consists of the parallel combination of a number of diode pump circuits with staggered time constants. Cooke-Yarborough and Pulsford have published an admirable study,[3] both theoretical and experimental, of this circuit, but the results do not include expressions for, for example, transient response or bandwidth of the randomly fluctuating output signal. An analysis in depth for these properties seems to be quite difficult, though it would be of considerable value to the designer of period meters.

It is clear, however, that the behavior of the two circuits is similar, qualitatively, in two important and related respects: both circuits have faster transient response at higher count rates, and both circuits exhibit more rapid fluctuations in measured count rate when operating at higher count rates.

In the log-diode circuit the amplitude of the fluctuations in the voltage output remains constant as counting rate varies. At the same time the bandwidth of the frequency spectrum of these fluctuations remains proportional to the counting rate. In the multiple-pump circuit the amplitude of the fluctuations in voltage output decreases somewhat with increasing counting rate. At the same time the fluctuations become more rapid, since the short-time-constant pumps dominate at high counting rates.

A differentiating network used with either type of logarithmic count-rate circuit to measure period must per-

Vol.
Pt. 1
1960

Paper 60-511, recommended by the AIEE Nucleonic and Radiation Instrumentation Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Nuclear Engineering and Science Conference, Cleveland, Ohio, April 6–9, 1959. Manuscript submitted December 1, 1958; made available for printing February 9, 1960.

BRUCE B. BARROW is with the SHAPE Air Defense Technical Center, The Hague, Holland, on leave from the Hermes Electronics Company, Cambridge, Mass. ROBERT MAITLAND is with the General Electric Company, Cincinnati, Ohio.

Mr. Barrow's contribution to this paper was supported in part by the General Electric Company, through a consulting agreement, and in part by the Hermes Electronics Company (formerly Hycon Eastern). An earlier version of this paper was presented in Paris on September 17, 1958, at the Colloque International sur l'Electronique Nucléaire, which was sponsored by the Société des Radioélectriciens.

form a dual function: it must differentiate the low-frequency portion of the voltage output of the count-rate circuit, and it must suppress the high-frequency fluctuations or noise. Various differentiating networks and their relative effectiveness in this dual role are discussed in what follows.

First, the frequency-response functions of differentiating networks offering varying degrees of suppression of high frequencies are presented. To compare these on a fair basis, parameters are so adjusted that all have an equally fast transient response. Then the frequency-response functions of the various networks are compared. Here narrow bandwidth is the desired quality, for, among networks with equally good transient response, the one that most effectively filters out noise must be preferred.

Conclusions drawn from the frequency-response comparison, valid whether log-diode circuit or multiple-pump circuit is used in the period meter, are checked analytically for a log-diode period meter. Finally, the practical realization of these networks is discussed.

THE PERFECT DIFFERENTIATOR

The complex-frequency function of a perfect differentiator is

$$H_{id}(s)=s \tag{1}$$

where s represents complex frequency $(j\omega+\sigma)$. If $j\omega$ is substituted for s in equation 1, the result expresses the fact that a perfect differentiator has a phase response of +90 degrees for all frequencies and an amplitude response that is proportional to frequency. Such a device is not physically realizable, for no circuit can maintain output proportional to frequency over an unbounded range of frequencies. On the other hand, for the reasons noted above, it is desirable that the differentiating network used in a period meter differentiate only low frequencies, and attenuate high frequencies. Such networks are physically realizable.

PRACTICAL DIFFERENTIATING NETWORKS

Consider the following functions representing realizable differentiating networks listed in increasing order of complexity:

$$H_1(s)=a_1s/(s+a_1) \tag{2}$$

$$H_2(s)=a_2^2s/(s+a_2)^2 \tag{3}$$

$$H_3(s)=(a_3^2+b_3^2)s/\{(s+a_3)^2+b_3^2\} \tag{4}$$

$$H_4(s)=(a_4^2+b_4^2)cs/(s+c)\{(s+a_4)^2+b_4^2\} \tag{5}$$

The a's, b's, and c are positive real constants. Each function, at low frequencies, is that of a differentiator. This may be seen by letting s approach zero. Except for the first, each offers attenuation at high frequencies, as may be seen by letting s increase without limit.

The function of equation 2 is that of the simplest realizable differentiating network, which may be built from one resistor and one capacitor; see Fig. 4(A). Physical realization of the other functions is discussed below. The differential equation corresponding to equation 3 is of second order, critically damped. Those corresponding to equations 4 and 5 are of second and third order, respectively, underdamped.

CHOICE OF PARAMETERS

To compare the different types of analog differentiators defined by equations 2 through 5, some arbitrary criteria had to be set up. First, it was decided to set the solution time at 15 seconds. (The solution time is defined as the time required for the output of the network to reach within 10% of steady-state value after sudden application of a unit ramp at the input. The unit-ramp function is zero for negative time and equal to t for positive time.) This is sufficient to fix a_1 and a_2. The unit-ramp response corresponding to equation 2 is

$$f_1(t)=1-exp(-a_1t) \tag{6}$$

and the constraint of 15-second solution time requires that

$$a_1=0.153/\text{sec (per second)} \tag{7}$$

The unit-ramp response corresponding to equation 3 is

$$f_2(t)=1-(1+a_2t)exp(-a_2t) \tag{8}$$

and the constraint of 15-second solution time sets

$$a_2=0.259/\text{sec} \tag{9}$$

Equation 4 contains two constants to be fixed. It is possible therefore to choose the ratio a_3/b_3 and then set the solution time equal to 15 seconds.

Solutions for various networks of this type showed that an optimum was reached when b_3 was set equal to a_3. The unit-ramp response corresponding to this condition is

$$f_3(t)=1+2^{1/2}exp(-a_3t)\sin(a_3t-3\pi/4) \tag{10}$$

This response shows an overshoot of only 4%. Although underdamped, the network is only slightly oscillatory. Setting the solution time at 15 seconds requires that

$$a_3=0.125/\text{sec} \tag{11}$$

The three constants of equation 5 were varied, relative to each other, to give a series of solutions, the best of which obtained when

$$a_4=b_4=c/2 \tag{12}$$

For this special case the unit-ramp response is

$$f_4(t)=1-exp(-2a_4t)-2exp(-a_4t)\sin(a_4t) \tag{13}$$

Setting the solution time at 15 seconds fixes

$$a_4=0.168/\text{sec} \tag{14}$$

The four ramp-response functions are shown, together, in Fig. 1. This figure gives a clear comparison of the transient response properties of the four types of networks. If for any application a solution time other than 15 seconds is desired, it is only necessary to multiply the above constants (the a's) by $15/x$, where x is the desired solution time, in seconds.

AMPLITUDE RESPONSE OF VARIOUS DIFFERENTIATING NETWORKS

The amplitude-response functions of the four types of networks are shown in Fig. 2. For each case the straight-line asymptotes are also given. The curves are obtained by setting $j2\pi f$ in place of s in each of equations 2, 3, 4, and 5, and by taking the absolute value of the result.

All four cases are seen to approach the response of the perfect differentiator, $2\pi f$, at low frequencies. The function of Case 1, $H_1(j2\pi f)$, derived from equation 2, offers no attenuation to high frequencies. The "breakpoint" frequency, at which the high-frequency and low-frequency asymptotes cross, is 0.024 c/s (counts per second).

The function of Case 2, derived from equation 3, attenuates high frequencies in inverse proportion to the frequency. This curve is symmetric about its maximum (0.0168 or −17.8 db), at a resonant frequency of 0.041 c/s.

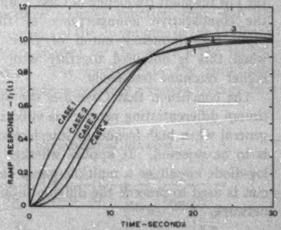


Fig. 1. **Transient-response functions of various differentiating networks with ramp input suddenly applied; parameters chosen for 15-second solution time**

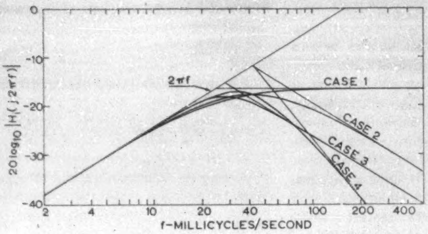


Fig. 2 (above). Frequency-response functions of various differentiating networks, together with straight-line asymptotes

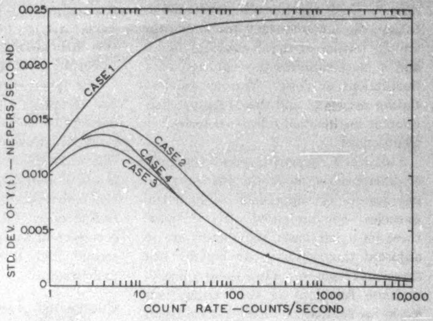


Fig. 3 (right). Standard deviation of indicated growth factor versus count rate, for log-diode period meters using various differentiating networks

Comparison with the function of Case 3, derived from equation 4, is interesting. The latter function is also symmetric about its maximum (0.0156) and this maximum is only slightly less than that of Case 2. It occurs, however, at 0.028 c/s, a much lower frequency. It is clear that the filter of Case 3 provides better discrimination than does the Case-2 filter against interfering signals at higher frequencies. The function of Case 4, derived from equation 5, on the other hand, resonates at a slightly higher frequency (0.034 c/s), and has a higher peak value. A filter of this type will, therefore, not be better than a Case-3 filter, even though the high-frequency portion of the response curve falls off faster, i.e., as $1/f^3$. Fig. 1 shows slightly less overshoot for the Case-4 filter, but the difference seems too small to justify the increase in circuit complexity.

When making comparisons from Fig. 2, it must be remembered that the logarithmic scales tend to give a distorted picture. The response of the various filters to wide-band noise would be more easily seen if the frequency-response curves were plotted with linear axes. Such a plot would show the advantage of the low resonant frequency of Case 3, and the comparative unimportance of the rapid high-frequency cutoff of Case 4 when this is obtained together with a higher resonant frequency.

The conclusion that Case 3 is the optimum differentiating network is valid in general when high-frequency interference is to be rejected. It applies whether a log-diode circuit or a multiple-pump circuit is used to precede the differentiating network.

PERFORMANCE WITH LOG-DIODE CIRCUIT

The performance of the various differentiating networks when operated with a log-diode circuit can be judged by examining the output fluctuations in indicated period at various steady-state counting rates. To describe these fluctuations, due entirely to the random nature of the neutron-detection process, it has proved convenient to work in terms of the growth factor, a variable that is proportional to the reciprocal of the period. The magnitude of the fluctuations may be determined in each case by solving the integral

$$var\{\gamma\}=\int_0^\infty \frac{nQ^2}{\pi}\cdot\frac{1/C^2}{\omega^2+(nQ/AC)^2}\cdot\frac{|H_i(j\omega)|^2}{A^2}\,d\omega, \qquad (15)$$

where $var\{\gamma\}$ denotes the variance (mean-square fluctuation about the mean) of the measured growth factor, γ. The factor nQ^2/π is the power density spectrum of the input to the log-diode circuit, a series of short current pulses, each of charge Q, occurring randomly at an average rate of n pulses per second. The second of the three factors in the integrand is the squared magnitude of the frequency-response function of the log-diode circuit, or, rather, of its linear approximation, which varies with n. The last of the three factors involves $H_i(j\omega)$, the frequency-response function of a differentiating network, and A, the log-diode constant, which is needed for calibration. If A is expressed in volts/-neper, then γ, the growth factor, will be expressed in nepers/second, and will be equal to the period reciprocal. The whole integrand gives the power density spectrum of the output from the period meter (the measured growth factor), and integrating over-all frequency gives the variance. The derivation of equation 15 and its physical significance have been discussed in more detail in an earlier paper.[1]

The integral of equation 15 has been solved for each of the types of differentiating network considered above. $H_i(j\omega)$ is obtained by setting $j\omega$ for s in each of equations 2 through 5. For Case 1, the simple differentiating circuit of equation 2,

$$var_1\{\gamma\}=a_1^2QN/2AC(N+a_1) \qquad (16)$$

where for convenience the count-rate variable

$$N\equiv nQ/AC \qquad (17)$$

is introduced. For Case 2, a second-order differentiating network, critically damped, from equation 3,

$$var_2\{\gamma\}=a_2^2QN/4AC(N+a_2)^2 \qquad (18)$$

For a second-order differentiating network, underdamped, from equation 4,

$$var\{\gamma\}=(a_3^2+b_3^2)^2QN/4a_3AC\{(N+a_3)^2+b_3^2\} \qquad (19)$$

and this reduces, in Case 3, to

$$var_3\{\gamma\}=a_3^2QN/AC(N^2+2a_3N+2a_3^2) \qquad (20)$$

For a third-order differentiating network underdamped, from equation 5,

$$var\{\gamma\}=\frac{(a_4^2+b_4^2)^2c^2(N+2a_4+c)QN}{4a_4AC(N+c)\{(N+a_4)^2+b_4^2\}\{(c+a_4)^2+b_4^2\}} \qquad (21)$$

This expression reduces, in Case 4, to

$$var_4\{\gamma\}=2a_4^2QN(N+4a_4)/5AC(N+2a_4)\times(N^2+2a_4N+2a_4^2) \qquad (22)$$

Fig. 3 shows plots of the standard deviation (square root of variance) of $\gamma(t)$, the measured growth factor, as a function of count rate, for each of the above cases. Here AC/Q has been taken equal to 20, a practical value for a well-designed period meter.[2] Fig. 3 confirms, for the log-diode period meter, the deductions made from Fig. 2. The differentiating network of Case 3 is seen to be the best, both because the peak of the random interference is lower and occurs

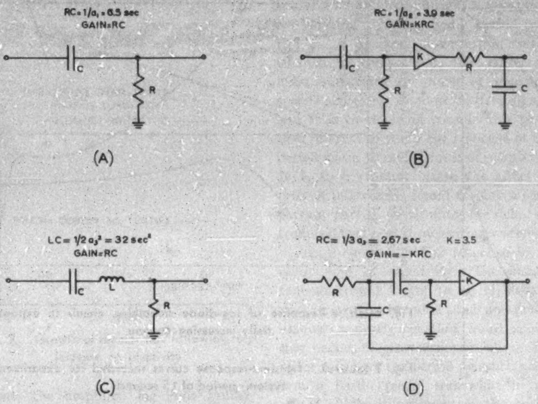


Fig. 4. Possible circuit configurations for various differentiating networks

at a lower counting rate, and because the interference at all counting rates is at least as low as for the other cases.

Realizability of Differentiating Networks

The above analysis is of use only if the mathematical functions of equations 2 through 5 can be realized in practical networks. Fig. 4 shows the essential elements of various differentiating networks having response functions discussed above. The differentiator gains listed are the outputs, in volts, for input ramps of 1 volt/sec. The numerical values shown for R-C (reactance capacitance) and L-C (inductance capacitance) products apply for a 15-sec solution time.

The simple differentiating circuit of Fig. 4(A) has the transfer function of equation 2. It is the simplest and most common differentiating network, and is of use wherever the filtering of high-frequency interference is not important.

The network of Fig. 4(B) realizes the function of equation 3. It can be built with no difficulty. The network of Fig. 4(C) would realize the function of equation 4, but the L-C product required to achieve the low resonant frequency (0.028 c/s) is hardly practical. Nor do mechanical filters offer an elegant solution, though one could certainly be built. An active R-C filter, on the other hand, is a simple solution. One such filter, which realizes the function of equation 4, is shown in Fig. 4(D). This network appears to have the disadvantage, when compared with that of Fig. 4(B), that drift in gain affects not only the calibration of the period meter, but also the shape of the frequency-response curve and the value of the resonant frequency. These effects are not important, however, in a circuit that is as highly damped as this one, and the difficulties with calibration remain the most serious. For example, a 25% increase in K would lower the resonant frequency of the circuit by 10% and would decrease the damping enough to give an overshoot on ramp response of 7%. Neither effect is very important, and the network would remain an effective differentiator and filter. The 25% error in calibration would, however, need to be corrected.

Many other active R-C filters could be designed to realize the function of equation 4. Kinariwala's interesting paper[4] includes a bibliography with important references.

Experimental Measurements of Transient Response

Experimental System

In order to observe the transient response of a physical period-measuring circuit, and to confirm, if possible, earlier conclusions regarding period meter design, an experimental period-measuring system was constructed on an analog computer. A log-diode circuit was used, followed by a differentiating network with a frequency function of the type given in equation 3. An exponential generator was used to provide various exponential inputs of current. A reactor simulator was also used, in place of the exponential generator, in order to study the response of the period-measuring system to ramp disturbances of reactivity.

The experimental system is illustrated schematically in Fig. 5. When used with a pulse input, from a neutron-detecting device, the log-diode circuit is preceded by a 50-pf (picofarad) "bucket" capacitor, aud the input pulses are 30 volts in amplitude. As a result, 0.0015 microcoulomb is dumped onto C for each neutron detection; this is the charge Q, introduced in equation 15. For the experimental study of transients, the bucket capacitor was by-passed, and a continuous-current input, $i(t)$, was provided to simulate the desired counting rate, according to the relation

$$i(t) = nQ \tag{23}$$

The current-voltage relationship of the log diode is given by

$$e_b = A \ln Bi_b \tag{24}$$

and the constant A, so defined, was equal to 0.1 volt/neper. Other system parameters are listed as follows:

$C = 0.5$ microfarad
$AC/Q = 33.3$
$R_1 = 1.33$ megohms
$C_1 = 3$ microfarads
$R_2 = 10$ megohms
$C_2 = 0.4$ microfarad

The two amplifiers shown in Fig. 5 were elements of the analog computer. They had extremely high gain (about 10^8), high input impedance, and negligible output impedance in this application. When connected as shown, each provided a voltage transfer function that was the quotient of the impedance in the feedback loop divided by the series input impedance. The amplifier in the count-rate meter thus provides a simple voltage gain

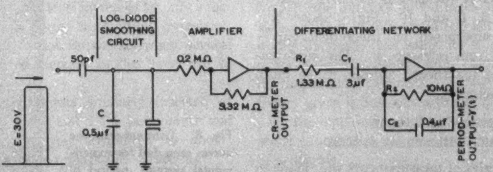
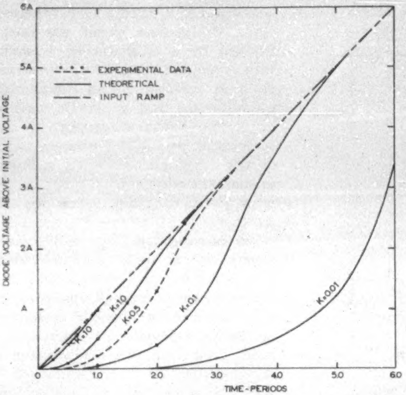


Fig. 5. Experimental system used for transient-response studies



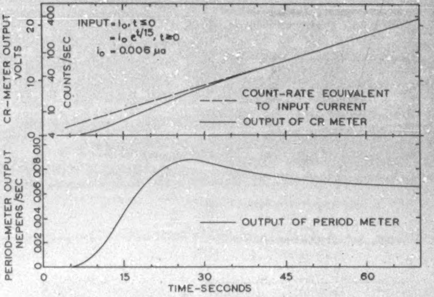


Fig. 6(left). Response of log-diode smoothing circuit to exponentially increasing currents

Fig. 7 (above). Transient-response curves recorded for experimental system; period of 15 seconds

of 46.6. The differentiating network has a transfer function that reduces to

$$E_{\text{out}}(s)/E_{\text{in}}(s)=R_2C_1s/(R_1C_1s+1)(R_2C_2s+1) \quad (25)$$

which is the expression of equation 3 with a numerator constant equal to 30 volts output per volt/sec input. The ramp response function of the differentiating network was measured and found to agree very closely with equation 8, with $a_2=0.25$/sec.

The growth-factor output, $\gamma(t)$, is a voltage, and may be displayed on a voltmeter or used as a control signal in a feedback-control loop. The diode voltage resulting from an input growth factor of one neper/sec (period of one second) is a ramp with a slope of 0.1 volt/sec. Therefore the output of the *CR* meter is a ramp with a slope of 4.66 volts/sec, and the output from the differentiating network is 140 volts, corresponding to a growth factor of one neper/sec. This is the required calibration constant.

TRANSIENTS IN THE LOG-DIODE CIRCUIT

Exponential-current inputs of the type

$$i(t)=i_0e^{t/T} \quad (26)$$

were impressed on the log-diode circuit. Measurements were made with the parameter K (defined as i_0T/CA) equal to 0.1, 0.5, 1, and 10. Fig. 6 shows test data superposed on theoretical curves taken from reference 2. The confirmation of the theory appears to be quite satisfactory.

TRANSIENTS IN ENTIRE PERIOD-MEASURING SYSTEM

A series of measurements was made to determine how well the entire period-measuring system would respond to exponential inputs, suddenly imposed. For example, a current input corresponding, by equation 23, to

$$n=4e^{t/15} \quad (27)$$

i.e., to a 15-second period (growth factor 0.067 neper/sec) initiated at 4 counts/sec, was imposed. The results of this test are shown in Fig. 7. One observes that the output transient, in measured growth factor, has decayed to within 10% of final value by the time n has reached 90 counts/sec. The overshoot in growth factor reaches a maximum of only 0.086 neper/sec (12-second period), far from saturation. This response indicates considerable improvement over period meters built previously.

In the experimental system, spurious noise limited the largest detectable period. Response to a 100-second period could, however, be measured and in this case only about 30 seconds were required to reach steady state.

To check short-period performance a current input corresponding to

$$n=4e^{t/3} \quad (28)$$

i.e., to a 3-second-period transient, was used.

The experimental result is shown in Fig. 8. The three recordings are, from top to bottom, the exponential current

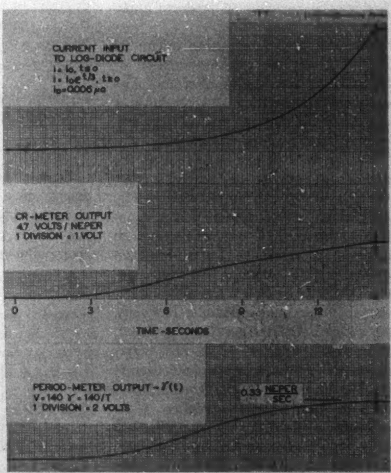


Fig. 8. Transient-response curves recorded for experimental system; period of 3 seconds

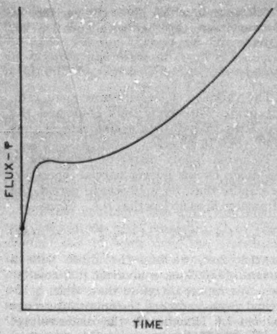


Fig. 9. Growth of reactor flux following step increase in reactivity

input, the amplified log-diode voltage (the parameter K equals 0.36 for this case), and the differentiating-network output (measured growth factor). The exponential generator was limited to about two decades, but measured growth factor reached 90% of steady state within these two decades, i.e., within 14 seconds. It is also clear from the measured curve that overshoot would be very small, less than it was with a 15-second period. This happens because the curve of diode voltage, though more distorted, approaches the straight-line asymptote more quickly (in only 9 seconds), and because the very distortion, with an excessive slope for a short time, shortens the differentiating-network transient. It is clear that periods somewhat shorter than 3 seconds could be measured accurately.

Although the effect of distortion in the diode-current transient can be beneficial, K should not be allowed to become less than unity, approximately, for periods of the order of the differentiating-network solution time. If, for example, a 12-second period is initiated from 1 count/sec, the form of the diode transient will be as above (K again equal to 0.36). From Fig. 6 it appears that a too steep slope in diode voltage will persist for longer than one period, i.e., for longer than 12 seconds. A much too high growth factor will thus be indicated by the period meter, and "scram" may be called for when all that was needed was patience. To avoid this type of response when operating with a reactor, a bias current, equivalent to two to four counts per second, should be kept on the diode.

Transients with Reactor-Simulator Input

Determining the response of the period measuring system to exponential inputs is useful for the basic insight it gives into the working of the system. Period control in a reactor, however, is often effected by control rods that either remain stationary or move at constant speed, according to some control signal, and it is therefore of considerable practical interest to study the response of the period meter to ramp inputs of reactivity, $\delta k/k$, to a reactor. Here the effect of prompt neutrons is found to play an important part in overcoming the delay introduced by a period-measuring system.

Fig. 9 shows the flux response to an input step of reactivity. The instantaneous growth factor ($\dot{\varphi}/\varphi$) is initially very large, then becomes small and finally assumes a nearly constant value as the flux begins to grow exponentially. The effect is to give the period-measuring system a hard "push" somewhat in advance of the appearance of the period that will shortly be present.

A series of measurements was made, accordingly, with the exponential-current generator replaced by a reactor simulator. For these tests the diode was fed a constant bias current of 0.006 μa, the equivalent of 4 counts/sec, to insure that transient response of the log-diode circuit would not be too sluggish. Fig. 10 shows test data simulating a start for a reactor just gone critical. Reactivity is inserted at 1%/minute for 17 seconds. Thus 0.28% $\delta k/k$ is inserted. Power level in the simulator, which is recorded also, eventually begins to increase with a stable 15-second period. In the third record the output of a fast period meter is shown, giving a nearly instantaneous record of the growth factor of power level. This meter lacks the time delays needed in a practical system to smooth out the noise from the neutron counting. According to this measurement, the "true" growth factor rises to a peak of about 0.13 neper/sec (period of 8 seconds) before settling down to a steady value of about 0.067 neper/sec (period of 15 seconds).

The measured growth factor, given last, rises to a peak of about 0.10 neper/sec, and requires about 20 seconds after the "true" growth factor has settled down to reach an accurate steady-state measurement. This is not alarming, however, for the power level increases by only a factor of four during the spurious overshoot, and this is not important when several decades of increase in flux level are to follow.

Reactivity was similarly inserted at rates of 0.5% and 1.5% per minute and the results were evaluated as above. There were no remarkable variations.

Finally, a test was run to see how quickly the period-measuring system would trip a 5-second-period alarm for various rates of shim-rod withdrawal. When reactivity was inserted at a rate of 6% per minute, the system signaled alarm after only 1.5 decades increase in power level.

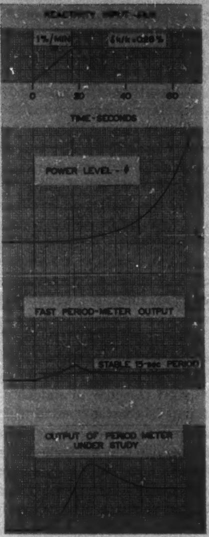


Fig. 10. Transient-response curves recorded for experimental system operated with reactor simulator

Conclusions

The foregoing analysis of linear differentiating networks rounds out the earlier work published on this subject, at least as applied to period-meter design. The curves given make it possible to judge whether, where electrical differentiation of a signal obscured by wide-band noise is required, the improved filtering performance of the slightly more complicated differentiating networks justifies their use. Although a quantitative analysis is given only for period meters using a log-diode circuit, the relative advantages of the various types of differentiating networks would be nearly the same for a period meter using a multiple-pump circuit for its logarithmic element. Though the theory given is basically a theory of analysis, the equations are simple enough to make practical a cut-and-try procedure of synthesis.

The experimental study was, in fact,

a check of just such a "synthesis" procedure, which had indicated that a log-diode circuit with AC/Q equal to about 30, followed by a differentiating network with a solution time of 15 seconds, would have adequate transient response for use in controlling startup on 15-second period. Transient response for a 3-second period and for longer periods was measured and found to give good confirmation of the theory, both for the log-diode circuit itself and for the entire period meter. It appears that periods as short as half a second would produce a satisfactory "scram" indication (*not* steady-state) if initiated at a count-rate lower than 100/sec.

The experiments with a reactor simulator are of particular interest, for they show that a period meter designed to have good response to suddenly imposed inputs of constant period will also respond well to the more complicated transient periods that are met when control rods are moved in a manner appropriate for an actual startup. The ability of the meter to detect short periods in a reactor is surprisingly good, because the prompt rise of reactor flux due to any disturbance of reactivity supplies a lead characteristic that tends to counter the time lags of the period-meter filters. Even when reactivity is inserted at a rate several times faster than normal, the "scram" indication provided by the period meter comes early enough for safety.

Appendix

The following example illustrates the use of the theoretical results in the analysis of a tentative design.

Two fixed-position fission chambers are cascaded, as shown in Fig. 11(A), to detect neutron density over a span of 8 decades. They are to be used in logarithmic count-rate and period meters designed to be effective over as much as possible of the 8-decade range. The proposed configuration is shown in Fig. 11(B). Each fission chamber is followed by a pulse-forming circuit and then by a log-diode circuit of the type discussed above. The diode voltage is amplified and fed to a meter, to indicate count rate, and also through a differentiating network to another meter, to indicate growth factor. When fission chamber 1 saturates, at 10^5 sec^{-1}, a relay operates to switch both the count-rate and period meters to the circuits associated with fission chamber 2.

The logarithmic element chosen for this application is a thermionic diode with a current-voltage characteristic that is logarithmic over about 5 decades, up to a current of 0.15 ma. Over this range the plate voltage changes by 0.10 volt per neper change in current (parameter A).

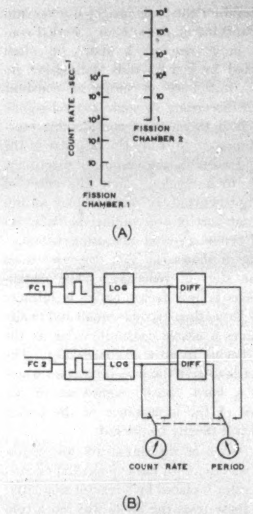


Fig. 11. Possible arrangement using fixed-position fission chambers cascaded to cover a wide range

Since the maximum diode current is 0.15 ma, and the maximum counting rate for the fission chamber is 10^5 sec^{-1}, the charge per pulse must be made equal to 0.0015 microcoulomb (parameter Q). This may be accomplished, as in Fig. 5, with 30-volt pulses and a 50-pf coupling capacitor.

Let us assume that the system is to be used, ordinarily, to help control startup on 15-second period, and that periods as short as 10 seconds shall be measured without perceptible transient during the transition from chamber 1 to chamber 2. This means that when an input of 10-second period is suddenly applied to the channel associated with chamber 2, the resulting output transient must decay before the count rate reaches 100/sec, i.e., before chamber 1 has saturated and switchover takes place.

In the design of the channel associated with chamber 2 the primary consideration must be fast transient response, for two reasons: 1. If transients too rapid to measure occur at higher flux levels they are potentially more dangerous than if they occur at low levels; 2. Noise at low count rates is not important because the output is connected to the indicating meter only at count rates greater than 100/sec. To insure fast response in the log-diode circuit we choose

$$AC/Q=20 \qquad (29)$$

This, with Q and A set as above, fixes the smoothing capacitor C, Fig. 5, as follows:

$$C=20Q/A=(20)0.0015/0.1=0.3\ \mu f \qquad (30)$$

To estimate the random fluctuation in the count-rate meter we use equation 33 of reference 2, which gives the variance of diode voltage, together with equation 29, above.

$$var\{e_b\}=AQ/2C=A^2Q/2AC=A^2/40 \qquad (31)$$

The variance of logarithm of count rate, not diode voltage, is thus 1/40, and the rms error in indicated count rate is thus $1/40^{1/2}$ or 0.16 neper.

To avoid highly distorted transients in the diode circuit during normal operation we require that, for a 10-second period, the parameter K be no less than 0.1. Since

$$K=i_0T/CA=n_0QT/CA=n_0/2 \qquad (32)$$

for this case, we bias the diode with a current of 0.003 μa, equivalent to 2 counts/sec. We expect therefore that, when a 10-second period reaches the count-rate range covered by chamber 2, the diode-voltage transient will be no more distorted than that shown in Fig. 6 for a K of 1.0. This transient has a too low slope for one period, then a too steep slope, and finally it settles down before 2.5 periods, or 25 seconds, have elapsed. Actually, with a bias current present, the transient is smoother than that for a K of 1.0, since under these conditions there is not a sudden step into a 10-second-period input, but the input follows a smooth transition, lasting over several periods, from the condition of no growth to that of exponential growth.

For a differentiating network we choose the configuration shown in Fig. 4(D), but with R-C decreased to 2 seconds. This has the effect of setting a_3 equal to 0.167/sec, and results in a solution time of only 11 seconds.

In predicting the transient performance of the period meter, we can make the assumption that steady-state is not reached at the output until a solution time of 11 seconds has elapsed after the diode-response curve of Fig. 6, for K equal to 1.0, has approached the ramp asymptote. Thus if a solution time of 11 seconds is added to the 25 seconds required by the diode circuit, a total elapsed time of 36 seconds, or 3.6 periods, is needed. This represents an exponential growth by a factor of 37, which implies that steady state would be reached by the time the count rate had reached 74 counts/sec. Thus, even a conservative calculation leads to the conclusion that the transient dies out before switchover at 100 counts/sec. The actual transient will be much shorter (see Figs. 7 and 8) because the differentiating networks will begin to respond before the transient in the log-diode circuit has died.

To get an idea of the magnitude of the random fluctuations to be expected in the neighborhood of the switchover point, we turn to equation 20, calculating for $N=2$ and for $N=5$ ($n=40$ and $n=100$, respectively). This gives

$$var_3\{\gamma\}_{n=40}=0.95\times10^{-4}$$

$$\text{standard deviation } (n=40)=0.0097 \text{ neper/sec}$$

$$var_3\{\gamma\}_{n=100}=0.43\times10^{-4}$$

$$\text{standard deviation } (n=100)=0.0066 \text{ neper/sec} \qquad (33)$$

Strictly, equation 20 holds only for

constant count rates, but, as stated in reference 2, when the count rate changes during the solution time of the differentiating network, the range of count rates may be examined to set bounds on the expected fluctuation of period-meter output. Thus we may estimate that the rms error of period-meter indication immediately after switchover will be between 0.007 and 0.009 neper/sec. Since the periods met in normal operation will be about 15 seconds (growth factor 0.07 neper/sec), this noise should cause no difficulties.

Different considerations affect the channel associated with chamber 1. This channel has to work over the transition between infinite period and 15-second period, and it has to be capable of indicating period at low counting rates, where noise is most important. Therefore lower noise than was needed with channel 2 must be obtained, even at the cost of slower transient response.

The parameters A and Q will be the same for both channels, however, since the former depends on the diode only and the latter depends on the diode and on the count rate at which saturation is to occur. The pulse-forming circuits will also be the same. There remain three system parameters to choose: value of the smoothing capacitor C, value of the diode bias current i_0, and solution time for the differentiating network. A proper choice depends on the reactor dynamics, rate at which reactivity is to be inserted, level of residual flux, etc. Without precise knowledge of these, we may however assume that the worst transient to be measured is no worse than a 15-second period beginning at 2 counts/sec. Thus we may set bias current as before and increase C to 0.45 μf, making AC/Q equal to 30. A still larger value of C might prove desirable, but might lead to overshoot because of the effect of the prompt neutrons; see Fig. 10. Similarly, the desire to provide more noise rejection without slowing response too much leads to a choice of the differentiating network of Fig. 4(D) as shown, with a 15-second solution time.

Calculation of response to various transients and of noise at various count rates may be made as indicated above for channel 2.

References

1. Random Fluctuations in Period-Meter Indications, W. M. Grim, Jr., B. B. Barrow, J. C. Simons, Jr. *Nuclear Science and Engineering*, New York, N. Y., vol. 1, 1956, pp. 80–91.

2. The Logarithmic-Diode Counting-Rate Meter and Period Meter, Bruce B. Barrow. *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 76, Sept. 1957, pp. 421–27.

3. An Accurate Logarithmic Counting-Rate Meter Covering a Wide Range, E. H. Cooke-Yarborough, E. W. Pulsford. *Proceedings*, Institution of Electrical Engineers, London, England, vol. 98, pt. II, 1951, pp. 196–203.

4. Synthesis of Active RC Networks, B. K. Kinariwala. *Bell System Technical Journal*, New York, N. Y., vol. 38, Sept. 1959, pp. 1269–1316.

A Digital Start-up Control for Air-Cooled Nuclear Reactors

SAMUEL N. LEHR
MEMBER AIEE

VERNON P. MATHIS
NONMEMBER AIEE

IN BRINGING a nuclear reactor up to power, the flux level goes through as many as 8 to 10 decades from the off position through full power. Somewhere within this range of power the reactor goes critical (effective multiplication factor equals one). This point where criticality is reached may be 7 decades below the full power point. In most stationary water-cooled reactors, this total spread is divided into three ranges of control. The start-up range normally covers 3 to 4 decades; the intermediate range, 4 to 5 decades; and the power range, 1 to 2 decades.

For a mobile reactor, it is advisable for simplicity to control the flux through only two ranges. These are the automatic start-up range covering about 6 or 7 decades and the power range covering 2 decades.

The rate of power increase in the start-up range is controlled by actuating control rods so that the period (time for the reactor power to increase by a factor of *e*) is always above the minimum safe value. If the reactor period decreased below this value, a start-up accident could occur and the reactor could run away.

There are many pieces of equipment now available to control the reactor through this range by analog methods. The outputs of various sensors are converted to electric voltages or currents and actuate relays, alarms, recorders, or meters which the operator continuously monitors. For an automatic start-up control, this method appears inefficient. The outputs of the sensors are already digital signals consisting of discrete pulses. The outputs of the controls are also digital in form—an on or off relaying action to the control rods. It appears that a method that would calculate and record the number of pulses and their time-variations (the reactor period) digitally will have an advantage over the analog system.

This paper describes a study made of such a system, the various system parameters required for a stable system, the safety functions necessary to prevent a nuclear accident, and a digital pulse-counting and period computation system.

It is desirable to review some of the principal reasons for preferring an over-all digital system for the start-up control. Primarily, digital circuit techniques result in a mode of operation of the active devices, i.e., transistors, that immensely increases the effective reliability of those devices. They are either off or on. Relatively large variations in gain, leakage current, or impedance level of the transistors have no effect on the ability of a properly designed circuit reliably to indicate yes or no. The operation is precise and positive and lends itself readily to many techniques of calculation. Digital methods, furthermore, may readily be adapted to exploit the precision and reliability of the input information to the fullest extent, thereby achieving the ultimate in system accuracy with no basic changes in system philosophy for improved models.

The preference for a digital display of level and period information is easy to understand. The analog meter reader concerns himself with less than 1% of the panel area occupied by his instrument. A digital indicator can provide much greater panel space efficiency and simultaneously simplify the problem of transferring indicated information to the operator's mind.

The use of a digital system will very likely result in an ultimate reduction of the logic requirements involved in the safety and control functions. Digital techniques incorporate, within the computation circuits, on-off processes that verify or correspond to many of the signals required in present malfunction networks. The very operation of the digital computer may be made to be, in some respects,

---

Paper 60-512, recommended by the AIEE Nucleonic and Radiation Instrumentation Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Nuclear Engineering & Science Conference, Cleveland, Ohio, April 6–9,1959. Manuscript submitted August 4, 1959; made available for printing February 9, 1960.

Samuel N. Lehr is with the Space Technology Laboratories, Los Angeles, Calif. Vernon P. Mathis is with the General Electric Company, Syracuse, N. Y.

The authors wish to acknowledge the valuable contributions of H. W. Abbott, J. D. Schmidt, and J. J. Suran.

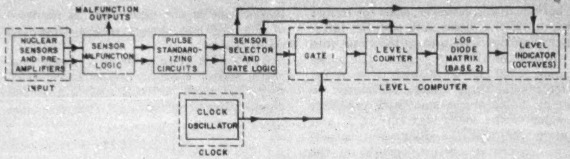

MALFUNCTION OUTPUTS
NUCLEAR SENSORS AND PRE-AMPLIFIERS
INPUT
SENSOR MALFUNCTION LOGIC
PULSE STANDARDIZING CIRCUITS
SENSOR SELECTOR AND GATE LOGIC
GATE 1
LEVEL COUNTER
LOG DIODE MATRIX (BASE 2)
LEVEL INDICATOR (OCTAVES)
LEVEL COMPUTER
CLOCK OSCILLATOR
CLOCK

Fig. 1. Digital level computer

a guarantee of the proper operation of the sensors.

The Count Rate Circuitry

Computing Scheme

The problem of level indication consists of determining the level of the reactor and providing means for display of this information. It is required to indicate the level of the reactor over a power range of 6 decades with an accuracy of at least half a decade. If this range is broken up into octaves and each of these is explicitly indicated, then the accuracy, to the nearest octave, will be within 1/3 of a decade. There will be 20 indications of interest, each corresponding to a doubling of reactor power with respect to the previous indication. A doubling indication, i.e., level indication using the radix 2 instead of *e*, has several notable advantages.

First, the concept of doubling lends itself readily to binary instrumentation as well as being an easy concept to understand from the operator's point of view. Second, the fortunate fact that there are approximately 10 octaves in 3 decades simplifies considerably the problem of display for the level indication since each nuclear sensor may be used to cover 10 of the 20 octaves. Assuming that this octave system is used, then the reactor level over the entire 6 decades may be indicated with sufficient precision by the decimal numers 0 to 19 corresponding exactly to the 20 octaves of reactor power in the start-up range.

Malfunction Signals and Pulse Standardization

Referring to Fig. 1, consider the sensor malfunction logic and the pulse standardizing circuits. These parts of the system are included to check the operation of the nuclear sensors, to discriminate the pulses to be counted, to supply necessary amplification of the signal, and to deliver uniform pulses to the digital circuits that follow. It is desirable to detect sensor malfunction with as little intervening circuitry as is possible to reduce the probability of malfunction detector failure.

The sensor malfunction logic is intended to check the output of each sensor for proper registration with the previous sensor output before information is accepted from it for delivery to the counters. A malfunction signal is capable of preventing the level counter signal from causing the sensor selector to switch to a new sensor. Proper operation signals for the sensors may be generated by noting that a certain count has been attained. A simple *RC* integrator operating on the raw input signal or a 2- or 3-stage magnetic core counter operating on the standardized pulses may be used to note the proper counting rate.

To generate standard pulses the signal may first be amplified to the extent necessary to furnish pulses of reasonable size from a differentiator. Since the signals arrive on a low impedance coaxial line, a common-base transistor may be used as a uniform gain voltage amplifier. This may be followed by emitter-follower stages to provide power gain.

Such a mode of operation minimizes the effects of transistor parameters on circuit gain. After differentiating the signal, to obtain a pulse for each fast rise time observed, the pulses are applied to a Schmitt trigger circuit. This includes the discrimination function if the input capacitor is chosen so that the circuit triggers only for pulses exceeding a prescribed energy level.

Sensor Selector and Gate Logic

In order to effect an over-all reduction in system complexity it is desirable that the two sensors share the use of as many of the start-up control circuits as possible. This implies ideally that a signal commutation scheme should be employed with respect to all of the circuits with the exception of the malfunction indicators. These should monitor each sensor at all times and indicate faulty operation as soon as it develops.

The sensor selector and gate logic network is incorporated into the system as early as possible to perform the switching or commutation function. It consists of several transistor gates and logic circuits operated by signals from the level counter and the malfunction logic block. The sensor selector is capable of passing the signal from the selected sensor into the level counter through Gate 1. Actually the chamber selector must also provide for passing a different signal from the second sensor to the period computation circuitry during the overlap interval when both sensors are in use. This operation will require a few additional logic elements.

Fig. 2 illustrates one commutation scheme. The gates, one on each sensor line, allow no signal to pass through unless both gate control lines have the proper signal. Assume that the 2-stage counter is always set to the state 00 when start-up begins. Then the gate on line 1 is the only one that allows pulses to enter the level counter. The level counter continues to accept pulses from sensor 1 until a predetermined count rate is reached, e.g., 131,072 pps (pulses per second).

The attainment of this value of count generates a pulse that is fed to the 2-stage counter, changing its state from 00 to 01. The gate on line 2 now is activated. The level counter also generates signals for the low count, e.g., 64 pps. Then if the reactor level should decrease so that the count falls below 64 pps, a "reverse-one-unit" signal is applied to the 2-stage counter. This switches in the lower sensor line. Finally, the condition 10 of the 2-stage counter activates a third gate that supplies a signal to the effect that the power range has been attained.

Vol.
Pt. 1
196

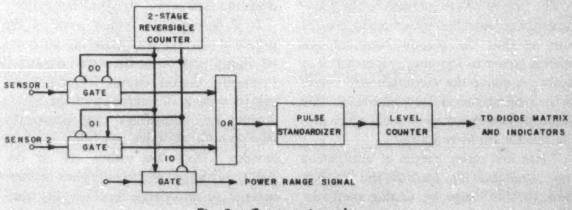


Fig. 2. Commutation scheme

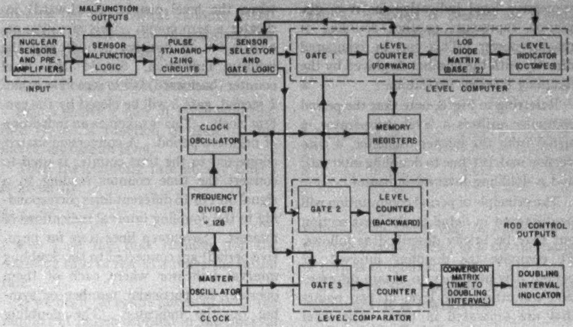


Fig. 3. Digital level and period computer

LEVEL COMPUTER

The level computer, shown in Fig. 1, operates as follows. Assume that the sensor selector allows a signal from sensor 1 to pass to gate 1. Gate 1 is operated by the clock oscillator at a repetition rate of 1 second, for example. The duty time is arranged so that gate 1 is open, allowing pulses to enter the level counter, for practically the entire second. A time of about 10 microseconds is ample for clearing the level counter at the end of each clock interval. The gate is therefore open almost all of the time and serves to mark the end of each second so that the reactor level, in terms of pulses per unit time, may be read out.

The log matrix converts the reading of the 17-stage level counter into a signal on one of 10 possible output lines corresponding to 10 successive doublings of counting rate, i.e., 10 different ranges of readings of the level counter. Thus a particular output signal, corresponding to one particular range of input pulse repetition rate over the sampling period is fed to the level indicator. Here, each of the 10 lines is capable of activating a particular number on a decimal digital-display device.

For example, if the reading of the level counter happens to lie between 1,024 and 2,047 counts when the gate closes at the end of some particular second, the log matrix will activate line 4 corresponding to the decimal digit "3." The number "3" would be displayed as the level indication and would be retained by suitable electrical or mechanical memory techniques until renewed or changed at the end of the next second. This number is interpreted as an indication of a particular range of neutron flux.

Suitable counting circuits employing transistor flip-flops may be designed in a straightforward manner for the frequencies of interest here. They are tolerant of transistor parameters and highly reliable.

The diode matrix is properly connected to the counter flip-flops to take the logarithm of the level count. The matrix is arranged primarily to supply a signal that lights the next indicator bulb the first time the next higher order flip-flop operates. Since the operation of flip-flop corresponds to a multiplication by 2, a logarithmic level indication results. Approximately 50 diodes are required for this matrix in an actual system.

It is seen that the level counter and log matrix are capable of activating 10 decimal digits corresponding to 10 octaves (3 decades) of reactor power. The entire range to be covered is 20 octaves. Therefore a "tens" column is required for the complete indicator as well as a "units" column. This may be done by utilizing signals from the sensor selector circuit to activate the required digits in the tens column.

For example, the tens indicator will be arranged so that if sensor 1 is being gated into the level counter, no tens digit will appear and the over-all indicator can show only the digits 0 to 9 in the ones column, corresponding to the first 10 octaves of power. When sensor 2 is gated in, a signal is fed to the indicator that displays a "1" in the tens column of the indicator. This, together with the ones column, gives octaves 10 to 19.

Note that there is a feedback path from the level counter to the sensor selector and gate logic block. When the level counter fills in the sampling interval to a predetermined value a signal is generated that can switch the sensor selector to the higher sensor. This switching operation is arranged so that it occurs if and only if the next sensor is operating properly as evidenced by the state of the sensor malfunction logic circuits.

The sensors are calibrated so that the high count of sensor 1 registers properly with the low count of sensor 2 and the high count of sensor 2 registers with the power range sensor; hence the registration of a particular count within a decade remains the same from sensor to sensor.

The counter readout circuits and the sensor selector circuit operate in either direction. That is, accurate level indications are produced whether the counting rate is increasing or decreasing. Therefore, the actual power level is always displayed and the display is revised each second.

The Period Circuitry

PERIOD COMPUTER

The problem of reactor period computation is fundamentally concerned with measuring the rate of change of reactor level. It is required to indicate the period of the reactor over a positive period range of, say, 3 to 30 seconds. Indications of specific selected values between these two extremes are also required as well as specific indications of positive period, negative period, and infinite period.

The accuracy of the measurements must increase as the period becomes shorter to a maximum accuracy of 5% for the very short periods. The delay time inherent in the instrumentation before a change in period is indicated should be small when compared to the time values of the periods of interest.

The period of the reactor at any power level is defined as the time required for the reactor power level to increase by a factor of e, i.e., 2.718. A considerable simplification in a binary digital system results if the factor "2" is employed as a multiplier rather than e, because the doubling operation is explicit in a binary system. For example, in a counter, the triggering of each successive flip-flop corresponds to a doubling of the count, i.e., a multiplication by 2. It is convenient, then, to think in terms of the doubling interval rather than the period.

The required values of period are easily converted to corresponding values of doubling interval. In addition, it is desirable that the "period" indication actually be displayed to the operator in terms of the doubling interval. This allows the operator, if he so desires, to check on whether the period computer is operating properly by simply timing the interval between level changes since each successive level indication, as it is displayed, represents another doubling of reactor power. This can be compared directly with the number displayed by the doubling interval indicator.

Referring to Fig. 3, note that the period computer utilizes a level comparator, a signal from the master oscillator, a conversion matrix (time to doubling interval), and a doubling interval indicator.

The principle of period computation will be discussed in detail in the next section but may be briefly described as follows. To compute the doubling interval, or period, it is necessary to obtain information concerning the level at two points that are separated in time. The level counter (forward) supplies an initial value of level at the end of each second, timed by the clock oscillator. It will be recalled that the initial level count was determined by adding, one by one, each of the pulses supplied by the sensor; the level being represented by the total number of counts obtained in one clock interval. To obtain information necessary to compute the doubling interval the pulses arriving from the sensor at the second point in time are used to reduce the level counter contents to zero by subtracting one count for each input pulse.

This count-back operation corresponds to using the level counter as a level comparator. The time required to empty the level counter at the new input pulse rate is compared to the fixed time required to fill the level counter initially in the forward direction. The time required to zero the counter by this method is related to the period of the reactor. For example, suppose that the period is infinite; that is, that the level has remained constant. In this case the level counter, having reached its initial count in just one clock interval, would be counted backward to zero in just the same amount of time. Considering the clock to be operating at 1 second for definiteness, a count-back time of 1 second would then always correspond to a period of infinity regardless of the value of the initial count. If the counter fails to count back to zero in 1 second or less, at the new pulse rate, the period is negative; that is, the level, as evidenced by counting rate, must have decreased. For all values of count-back time between zero and 1 second the period is positive, each value of time corresponding to a particular period regardless of level.

In the system shown in Fig. 3, the master oscillator supplies pulses to the time counter of the level comparator whenever gate 3 is open. The gate is opened by the clock oscillator and closed when the level counter (backward) indicates a zero count. This circuitry accounts for all positive periods and the infinity indication. Should the level counter (backward) fail to zero in less than 1 second, gate 3 will be closed by the system clock. This is taken as an indication of negative period. A conversion matrix connected to the time counter is used to convert the time counter reading to a signal on 1 of 16 different lines corresponding to the doubling interval indications of interest. Seventeen lines (one for negative period) are connected to the doubling interval indicator where each of them operates its particular number or symbol on the indicator. The doubling interval indicator is capable of holding each indication until the next clock cycle when the display is changed or renewed. The display may be similar to the level display, if desired, or of a different form. Five values of doubling interval are chosen to activate the five levels of operation and safety of the reactor. Rod control outputs are produced directly from these five selected indicator lines.

The function of the memory registers is to reduce the effect of the statistical variations of the nuclear sensor pulse rate and to increase the precision of the time counter indication. It is desirable to delay the count-back operation for as long a time as possible so that the time required to zero the level comparator counter will be appreciably different from 1 second even for large values of doubling interval. On the other hand, it is required that the doubling interval indications recur frequently as compared to the shortest doubling intervals of interest.

The memory registers are used to achieve this dual result in the following manner. Assume that the forward level counter has just reached its reading and that the level indicator has been triggered by the clock, giving up-to-the-second information on reactor level. The contents of each flip-flop of the counter are quickly shifted at this time to the first memory register and held there for one clock interval. The counter meanwhile is being refilled with new level information. This in turn is fed to the first memory register at the end of the next clock interval while the contents of register 1 are shifted to the second memory register.

Finally level information is shifted into the level comparator counter where it is counted back at the current counting rate. This results in a new doubling indication each second, based on what the reactor level was in that last second, but compared with level information obtained several seconds ago. System considerations will dictate an optimum number of these 17-bit memory registers. Two such registers, each consisting of 17 magnetic core elements, seem to offer a maximum advantage.

The problem of avoiding a delay while the memory registers fill when switching to the next sensor is solved by allowing a small amount of sensor overlap. Thus gate 2 continues to be fed from sensor 1, for example, as long as the level comparator counter is being supplied with sensor 1 information from the memory registers. The line feeding gate 1 switches to sensor 2 when the level counter so indicates and level information based on sensor 2 begins to move down the memory registers toward the level counter (backward) in the comparator. At the time that the comparator is to receive its first charge of in-

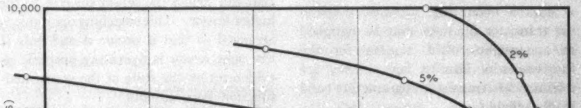


Fig. 4. Probability, sample size, and per-cent accuracy

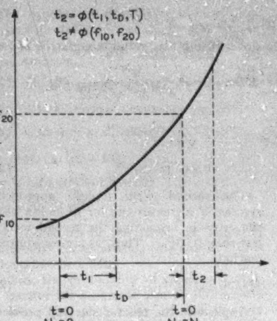


Fig. 5. Pulse counting rate versus time

formation from sensor 2, the line feeding gate 2 switches to sensor 2. Thus the countback always occurs at a rate determined by the sensor that supplied the count then present in the level comparator counter.

The fact that the sensor input information is statistical in nature means that the level counter will not read exactly the same value each time even for a fixed reactor level. Curves showing the total count necessary to produce a desired probability of an individual answer that is correct to a prescribed per cent are shown in Fig. 4.

The conclusion to be drawn from these curves is that the pulse information will be sufficiently precise for the level indication and the resulting period computation. For example, if the total count is 1,000, then the probability that any individual level count will be in error by more than 5% is only 0.1. Thus a clock interval of 1 second could be used, if desired, and a counter with enough stages to count to about 200,000 would be required to accommodate the maximum input pulse repetition rate.

Since the control rods are driven by outputs from the doubling interval indicator, it is desirable to include an extra feature in this indicator circuit to be certain that it does not respond to occasional drastic random variations in sensor pulse rate. To accomplish this a digital statistical weighting circuit may be included in the doubling interval indicator. The reactor period is known to vary smoothly and relatively slowly. The weighting circuit takes advantage of this by incorporating transistor and diode logic networks that prevent the period indicator from changing more than plus or minus one doubling interval increment per second. This technique effectively smooths out any radical random variations that remain in the information after the previous processing.

PERIOD COMPUTATION SCHEME

Fig. 5 shows a plot of pulse counting rate (f) as a function of time (t) for increasing reactor level.

Then the counting rate at any time t is given by

$$f=f_{10}e^{t/T} \tag{1}$$

where T is the time required for the counting rate to rise to e times the initial rate.

If we begin to count pulses at $t=0$, the number of pulses counted in the interval t_1, is:

$$N=\int_0^{t_1} f_{10}e^{t/T}dt \tag{2}$$

$$N=f_{10}T(e^{t_1T}-1) \tag{3}$$

If we wish to count this same number of pulses (N) after a certain delay time, t_D, the time interval, t_2, required to count N pulses may be determined by:

$$t_2=\frac{N}{\bar{f}_2} \tag{4}$$

where $\bar{f}_2$ is the average counting rate in the interval t_2.

$$\bar{f}_2=\frac{1}{t_2}\int_0^{t_2} f_{20}e^{t/T}dt \tag{5}$$

$$\bar{f}_2=\frac{f_{20}T}{t_2}(e^{t_2/T}-1) \tag{6}$$

Substituting equations 3 and 6 in equation 4, we have:

$$t_2=\frac{f_{10}T(e^{t_1/T}-1)}{f_{20}T(e^{t_2/T}-1)}\cdot t_2 \tag{7}$$

Hence

$$\frac{f_{20}}{f_{10}}=\frac{e^{t_1/T}-1}{e^{t_2/T}-1} \tag{8}$$

But, from equation 1, we have

$$f_{20}=f_{10}e^{t_D/T} \tag{9}$$

$$\frac{f_{20}}{f_{10}}=e^{t_D/T} \tag{10}$$

Combining equations 8 and 10:

$$e^{t_D/T}=\frac{e^{t_1/T}-1}{e^{t_2/T}-1} \tag{11}$$

Solving for t_2:

$$t_2=T\ln[e^{-t_D/T}(e^{t_1/T}-1)+1] \tag{12}$$

This is the basic relation of interest in computing the period. It states that for any given period and time delay there is one and only one corresponding value of t_2. Thus, a measurement of t_2 uniquely determines the period. The choice of values for t_D and t_1 may be optimized to produce a large spread in t_2 for the range of periods of interest.

Table I. t_2 Spread in Seconds for Various Values of t_1 and K

K	t_1 0.64	 1.00	 1.28
1	0.11	0.23	0.35
2	0.18	0.39	0.57
3	0.25	0.51	0.72*
4	0.31	0.60	0.82
5	0.34	0.64	0.86

To solve the equation by computer techniques, it is rewritten as

$$t_2=T\ln[e^{-Kt_1/T}(e^{t_1/T}-1)+1] \tag{13}$$

where

$$kt_1=t_D$$

Multiplying and factoring gives

$$t_2=T\ln[e^{-(K-1)t_1/T}(1-e^{-t_1/T})+1] \tag{14}$$

In the form of equation 14, t_2 was computed for values of T ranging from 3 to 30 seconds, corresponding to integral values of doubling interval ranging from 2 to 20 seconds; K from 1 to 5; and t_1 for the three values: 0.64 second, 1 second, and 1.28 seconds.

A measure of the spread in t_2 obtained by utilizing each of the above combinations of t_1 and K is the difference in countdown time between that corresponding to a doubling interval of 20 seconds and the value corresponding to 2 seconds.

The results obtained are tabulated in Table I, which shows that longer values of counting interval (t_1) and larger values of delay time (K) results in a larger time spread in t_2. Since the values of t_2 are used to compute the doubling interval, a greater spread in t_2 can contribute to more accurate indications of doubling interval. A value of $t_1=1.28$ seconds and $K=3$ are considered to be optimum for the present system; this is indicated in Table I by asterisk. t_1 cannot be chosen too large for it represents the elapsed time between readouts; K cannot be too large since it corresponds to the length of the memory and is related to the required amount of sensor overlap. The particular value 1.28 seconds was selected because it can be counted by just filling a 7-stage time counter using a master oscillator of 100 cps (cycles per second).

It should be noted that the derivation of equation 12 was in terms of the period, T, defined as the time required for the level to reach e times its initial value. Table I is presented in terms of the doubling interval, D, defined as the time required for the level to double. The explicit equation corresponding to equation 12 for t_2 in terms of D is:

$$t_2 = \frac{D}{\ln 2} \cdot \ln[2^{-t_D/D}(2^{t_1/D}-1)+1] \tag{15}$$

The entire derivation thus far assumes only that the pulses are increasing exponentially with a fixed period. The assumption of constant period is justified by considerations of the physical system to which the analysis is to be applied. The problem of the statistical variation in the pulse rate is considered in the Appendix. This serves, however, only to impose on the foregoing analysis a certain probability of obtaining an answer correct to the desired accuracy.

Conclusions

Some important conclusions arrived at after a careful consideration of the results of the study are noted below:

1. An over-all digital system, using solid-state devices, for performing the functions of the start-up control is definitely feasible. This includes an all-digital computation technique for level and for period, a digital display of level and period, and the replacement of relay logic by semiconductor logic.
2. The bulk of the problem in applying solid-state digital techniques to the control lies in the level and period computation problems. The amplifiers, pulse shapers, and logic circuits can be approached with the aid of circuit techniques already in use. It is noted, however, that perhaps half of the total number of circuit components are required just to duplicate exactly the present safety and control interlock circuits.
3. The over-all digital system does not require many more components than the analog system. The difference becomes insignificant if digital display is desired regardless of the computation scheme employed.
4. The use of solid-state devices in the start-up control will result in great size, weight, and power reduction and an increase in reliability. Important extra reliability, thermal insensitivity, and component interchangeability will result from the use of an all-digital system.

Appendix

Statistical Considerations

The use of the Poisson distribution to describe the statistics of nuclear events has been justified by both theory and experiment.[1,2] This function may be written as:

$$f_p = \frac{M^n e^{-M}}{n!}$$

where

f_P = the value of the probability for a given M and n

M = the arithmetic mean, or average count for a large number of trials

n = the particular number, or count, observed in one trial

Feller[1] states that "For any fixed M, the simplest physical assumptions lead to the Poisson distribution as defined above as the probability of finding exactly n events within a fixed interval of specified length." This corresponds exactly to the problem in the present study.

Pollard[2] justifies the use of the Poisson distribution by quoting the experimental agreement of count data with the results predicted by the theory. He goes on to point out that a fluctuation of count "as great as the square root of the number of particles counted in any one observation" can be expected and remarks that a relation between probability, sample size, and accuracy exists.

The theory developed in this Appendix verifies Pollard's statements and results in useful quantitative data regarding the inter-relations of probability, sample size, and accuracy of measurement.

Mean Deviation (From the Mean) of a Poisson Distribution

The Mean Deviation (from the mean) for a Poisson distribution is of considerable interest because of its rather simple relation to the arithmetic mean.

The Mean Deviation (MD) is defined as follows[3]:

$$MD = \sum_n f_i \frac{(n-M)}{N} \tag{16}$$

Where M is the arithmetic mean, N is the total weight or probability, and f_i is the value of the probability n units from the arithmetic mean.

For a Poisson distribution f_i is given by the following relation.

$$f_P = \frac{M^n e^{-M}}{n!} \tag{17}$$

Substituting this value of f_i in equation 16:

$$MD = \sum_n \frac{M^n e^{-M}}{n!} \frac{(n-M)}{N} \tag{18}$$

Making the approximation that the distribution is symmetrical about the arithmetic mean, it is convenient for purposes of computation to consider only those values of n less than M. Consequently N, the total probability, will be 0.5 and equation 18 becomes

$$MD = 2e^{-M}\left(\sum_{n<m} \frac{M^n}{(n-1)!} - \sum_{n<m} \frac{M^{n+1}}{n!}\right) \tag{19}$$

$$MD = -2e^{-M}\left(\frac{M^{M+1}}{M!}+1\right) \tag{20}$$

An approximation to $M!$ is given by Sterling's equation[4]:

$$M! = e^{-M} M^M \sqrt{2\pi M} \tag{21}$$

Substituting this value in equation 20,

$$MD \simeq -2e^{-M} \frac{M^{M+1}}{e^{-M} M^M \sqrt{2\pi M}} - 2e^{-M} \tag{22}$$

$$\simeq -2\frac{M}{\sqrt{2\pi M}} - 2e^{-M} \tag{23}$$

$$\simeq -0.8\sqrt{M} - 2e^{-M} \tag{24}$$

The second term rapidly approaches zero as M increases and for M as low as 10 the error in neglecting the second term is less than 0.03%. Thus, as a close approximation:

$$MD \simeq 0.8\sqrt{M} \tag{25}$$

Applied to the reactor start-up problem, this means that for nominal mean counts of 10 pulses per second, 100, 1,000, 10,000, and 100,000 we may expect, respectively, average deviations from the mean of approximately 3, 10, 30, 100, and 300 counts. Thus the expected per-cent error goes down as the number of counts increases.

Probability, Sample Size, and Per-Cent Accuracy

The interrelations of probability, sample size, and accuracy are implicit in the Poisson distribution. It is not convenient however, to calculate the values leading to design curves by using the equation directly.

Waugh[5] presents a table entitled "Probability that an Item Chosen at Random from a Normal Distribution Will Lie at Least as Far From the Arithmetic Mean as the Stated Number of Standard Deviations." The table lists probabilities in terms of x/σ where x is the item chosen and σ is the standard deviation.

Before making use of this table it is necessary first to define a "standard deviation" for the Poisson distribution.

The Poisson distribution is given by

$$f_P = \frac{M^n e^{-M}}{n!} \tag{26}$$

and a normal distribution is given by

$$f_N = \frac{1}{\sigma\sqrt{2\pi}} \cdot e \frac{-(n-M)^2}{2\sigma^2} \tag{27}$$

Assuming that the curves have a similar shape in the region of interest we may equate 26 and 27, giving

$$\frac{M^n e^{-M}}{n!} = \frac{1}{\sigma\sqrt{2\pi}} \cdot e \frac{-(n-M)^2}{2\sigma^2} \tag{28}$$

Since we are particularly interested in the region near $n = M$, and since the result must hold for $n = M$, we may substitute n for M in equation 28. This gives

$$\frac{M^M e^{-M}}{M} = \frac{1}{\sigma\sqrt{2\pi}} \tag{29}$$

Applying Sterling's equation for $M!$,

$$\frac{M^M}{M^M \sqrt{2\pi M}} e^{-M} e^M = \frac{1}{\sigma\sqrt{2\alpha}} \tag{30}$$

And, simplifying,

$$\sigma \simeq \sqrt{M} \tag{31}$$

which is the required result.

Vol.
Pt. 1
196

The process of computing the relation between probability, accuracy, and sample size by Waugh's table may best be shown by an example.

1. Select a sample size, M, e.g., 100.
2. Select a desired accuracy, e.g., 20%.
3. Compute x, σ, and x/σ.

$\sigma=\sqrt{M}$, $(\sqrt{100}=10)$

$X=\%(M)$, $(0.2\times100=20)$

$X/\sigma=20/10=2$

4. Look up in the table the value of probability corresponding to $X/\sigma=2$. This is found to be 0.0454.

Thus, the probability of an error of 20% or greater for a sample size of 100 is only 0.0454. The probability of the correct answer, within 20%, is 0.9546—over 95% certain for each individual calculation.

This computation when repeated for other values of sample size and % accuracy leads to the family of curves shown in Fig. 4. These are used to anticipate computation accuracy and to determine required sample size for the start-up control problem.

The period error for a single calculation is estimated as 5% for a doubling interval of 2 seconds whenever the probability of the correct level indications exceeds 0.9.

Additional material in this area can be found in the list of references.

References

1. An Introduction to Probability Theory and its Applications, Vol. I (book), **William Feller.** John Wiley & Sons, Inc., New York, N. Y., 1950.

2. Applied Nuclear Physics (book), **E. Pollard, W. L. Davidson.** John Wiley & Sons, Inc., 1942.

3. Mathematical Tables from Handbook of Chemistry and Physics (book), C. D. Hodgman, editor. Chemical Rubber Publishing Company, Brooklyn, N. Y., 1947.

4. The Elements of Probability Theory (book), **H. Cramer.** John Wiley & Sons, Inc., 1955.

5. Statistical Tables and Problems (book), **A. E. Waugh.** McGraw-Hill Book Company, Inc., New York, N. Y., 1952.

6. Safety Systems for Nuclear Power Reactors, **I. M. Jacobs.** *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 76, Nov. 1957, pp. 670–73.

7. An Electroluminescent Digital Indicator with Translation Logic, **E. A. Sack.** *Ibid.*, vol. 77, Mar. 1958, pp. 113–18.

8. Transistor Reversible Binary Counter, **R. L. Trent.** *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 49, 1952, p. 1562.

9. Reversible Binary Counter, **Fenmore.** *Electronic Engineering*, London, England, vol 27, 1955, p. 204.

10. Pulse and Digital Circuits (book), **J. Millman, H. Taub.** McGraw-Hill Book Company, Inc., 1956.

11. Principles of Transistor Circuits (book), **R. F. Shea,** et al. John Wiley & Sons, Inc., 1953.

12. Transistor Circuit Engineering (book), **R. F. Shea,** editor. John Wiley & Sons, Inc., 1957.

A High-Speed Microcurrent Amplifier

JOHN A. DEVER
ASSOCIATE MEMBER AIEE

LOUIS SICKLES
NONMEMBER AIEE

BY THE USE of feedback amplifier techniques considerable improvement has been made in the speed and/or stability of the familiar electrometer tube micromicroammeter. Simplicity and compactness have been retained. With 100 feet of input cable, the amplifier has a 150-millisecond time constant at 10^{-11} amperes full scale with 0.05% per day stability.

Control systems for nuclear reactors frequently call for operation of linear and logarithmic current measuring instruments down to levels of 10^{-11} amperes and below. This is especially true of those for research reactors and critical facilities. Amplifiers designed for these applications should be stable, insensitive to system disturbances, reliable and fast even with long signal cables. Good speed of response provides rapid indication and insures that safety functions are not compromised. A related consideration is the amplifier transfer function; bandwidth limitations can markedly affect the performance of a controller in a closed loop system where the amplifier serves as the flux measuring device. Also, broader usage as a laboratory or research instrument is likely to require the widest bandwidth that can be obtained.

The micromicroammeter to be described is the result of a study to find the practical limits on speed for the familiar electrometer tube d-c amplifier. Since simplicity and modest cost were deemed important, the more complex stabilized amplifier has not been considered. Instead, a single-ended input stage has been adopted in which drift is limited by single or double regulation of all important d-c potentials. What drift remains is made acceptably small by the choice of 100% feedback and a 10-volt span.

The problem of high speed and good transient response in a micromicroammeter raises design difficulties not usually encountered in feedback amplifiers. This is chiefly because the input time constant, created by capacity from the input terminal to ground, can vary over a range of 10^9 for the 8-decade region of operation extending from 10^{-11} to 10^{-3} amperes. The effect of this factor in determining amplifier stability can conceivably be offset by altering attenuation networks for each current range. However, this can be awkward and expensive to do; when easy conversion to logarithmic measurement is desired absolute stability everywhere in the frequency spectrum is a necessity. With this in mind, appropriate care is taken in gain and phase manipulation to arrive at an optimum combination of over-all gain and bandwidth.

Closed Loop Design

Fig. 1 contains all the parameters needed to characterize the micromicroammeter. The forward gain block consists of three stages of gain and a cathode follower output. C_{IN} represents stray, wiring, and cable capacity, and capacity of the chamber or other source producing the current I_{IN}. A certain amount of capacity, C_F, exists from output to input, intentionally or otherwise, and is generally associated with the feedback resistor, R_F.

Degenerative feedback applied as show causes the output drift to be equal to the amplifier drift referred to the input, reduces the input impedance to $=R_F/A(s)$, and reduces the output impedance to $=1/g_M A(s)$ where g_M is the transconductance of the cathode follower output stage.

The closed loop transfer function is

$$\frac{e_0}{I_{in}}=-\frac{R_f}{\dfrac{sR_f(C_{in}+C_f)+1}{A(s)}+sR_fC_f+1} \qquad (1)$$

where

$s=j\omega$

As a first approximation, assume that $A(s)$ is large and constant and equal to A_o. In every case we will consider, C_{IN} is much larger than C_F. Then

---

Paper 60-513, recommended by the AIEE Nucleonic and Radiation Instrumentation Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Nuclear Engineering & Science Conference, Cleveland, Ohio, April 6–9, 1959. Manuscript submitted December 11, 1958; made available for printing February 9, 1960.

John A. Dever and Louis Sickles are with the Minneapolis-Honeywell Regulator Company, Philadelphia, Pa.

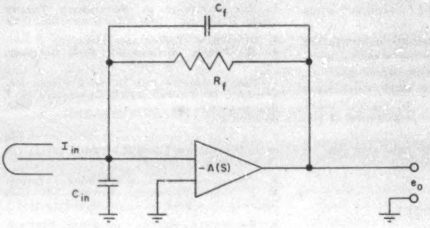


Fig. 1. Micromicroammeter system schematic

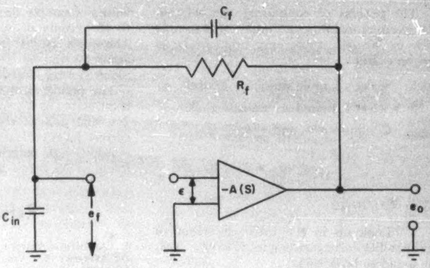


Fig. 2. Open-loop configuration of miromicroammeter

$$\frac{e_0}{I_{in}}=-\frac{R_f}{s\left(\frac{R_fC_{in}}{A_0}+R_fC_f\right)+1}=-\frac{R_f}{s\left(\frac{T_{in}}{A_0}+T_f\right)+1} \qquad (2)$$

where

$T_{IN}=R_FC_{IN}$

$T_F=R_FC_F$

$C_{IN}\gg C_F$

Obviously, as the gain gets very large, T_F becomes the limiting factor in obtaining wide bandwidth.

A somewhat more accurate description of the closed loop is possible by assuming $A(s)$ to be a simple lag, a fair approximation of the actual function.

$$A(s)=\frac{-A_0}{sT_a+1} \qquad (3)$$

Then

$$\frac{e_0}{I_{in}}=-\frac{R_f}{s^2\frac{T_{in}T_a}{A_0}+s\left(\frac{T_a+T_{in}+A_0T_f}{A_0}\right)+1} \qquad (4)$$

Examination of equations 2 and 4 leads to the following conclusions for the design of a wide band instrument.

1. Low-frequency forward gain, A_o, must be large to degenerate the input time constant.

2. C_F must be small on the lowest ranges in order that $T_F\leq T_{IN}/A_o$. On higher ranges the effect of T_F reduces because the gain $A(s)$ becomes less than A_o as the frequency increases.

3. The bandwidth of $A(s)$ must be as wide as possible, i.e., T_A must be small, for wide instrument bandwidth on the higher current ranges.

Open Loop Design

Equation 1 can be rewritten in the form

$$\frac{\theta_0}{\theta_i}=\frac{1}{\beta}\,\frac{A\beta}{1+A\beta}$$

giving

$$\frac{e_0}{I_{in}}=\frac{-R_f}{sT_f+1}\;\frac{\frac{sT_f+1}{sT_{in}+1}A(s)}{1+\frac{sT_f+1}{sT_{in}+1}(As)} \qquad (5)$$

Care must be exercised in identifying the relative importance of the various terms. The denominator $(1+A\beta)$, not β alone, controls the response in most instances. However, loop gain $A\beta$ is easily recognized and, referring to Fig. 2 for the physical interpretation, the open loop transfer function can be written

$$\frac{e_f}{\epsilon}=-A(s)\,\frac{sT_f+1}{sT_{in}+1} \qquad (6)$$

The time constant, T_{IN}, formed by the feedback resistor and the input capacity, is the dominant factor in determining the stability of the amplifier. This time constant is widely variable due to different cable lengths and range resistors and the corresponding break frequency can move throughout the frequency spectrum.

If the effects of both T_F and T_{IN} are taken into account, the examination of the open loop becomes complicated because both the coefficients are independently variable. For a basic analysis of the instrument, T_F can be assumed zero. Later, the proper variation of T_F can be calculated to get the maximum benefit from the lead effect.

Let T_F-0, then

$$\frac{e_f}{\epsilon}=-A(s)\,\frac{1}{sT_{in}+1} \qquad (7)$$

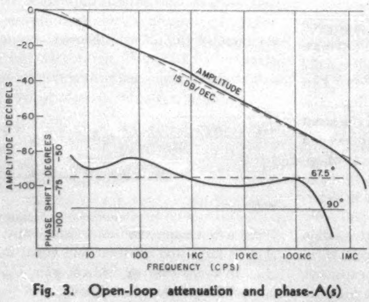


Fig. 3. Open-loop attenuation and phase-A(s)

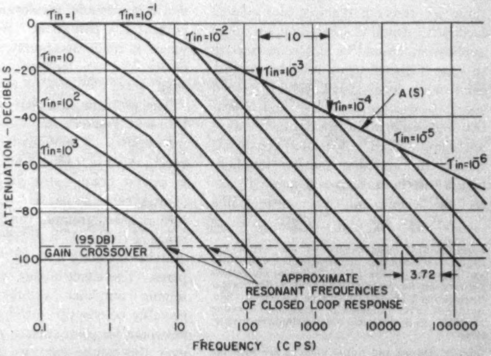


Fig. 4. Open-loop attenuation for various input time constants

Vol.
Pt. 1
1960

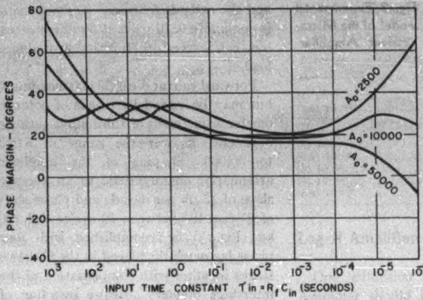


Fig. 5. Phase margin as a function of loop gain and input time constant

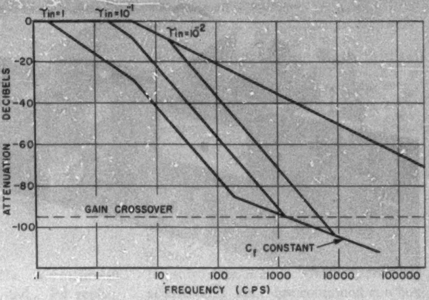


Fig. 6. Attenuation characteristic with C_F present

If no lead term exists, the phase shift of $A(s)$ plus $1/sT_{IN}+1)$ must not exceed 180° for all frequencies up to the frequency of unity gain to prevent oscillation. Actually, for acceptable transient response, 180° must not be approached too closely. Consequently, a phase margin of approximately 22.5° has been adopted. The function of $A(s)$ is shown in Fig. 3. Neglecting for the moment how this is obtained, notice that the average attenuation rate is 15 db (decibels) per decade, and the average phase shift is 67.5°.

Fig. 4 shows the lag term $(1/sT_{IN}+1)$ added to the attenuation curve of $A(s)$. The approximate resonant frequency of the closed loop function may be picked off the graph for a specific input time constant and loop gain.

For example, measuring 10^{-10} amperes full scale, a 10^{11}-ohm resistor is used which, with 1,000-$\mu\mu$f (micromicrofarad) input capacity, gives an input time constant of 100 seconds. If a d-c loop gain of 50,000 is employed, the approximate resonant frequency of the instrument is 20 cycles. If a gain of 1,000 is used, the resonant frequency is close to 1.4 cycles. Note that the resonant frequency of the instrument does not increase by the same factor as a reduction in the input time constant. If T_{IN} is changed by a factor of 10, for instance, by changing range with C_{IN} fixed, the resonant frequency changes by only a factor of 3.72.

Shown in Fig. 5 is the instrument phase margin for a specified loop gain neglecting any lead term in the feedback. This figure is constructed by adding the phase shifts of $A(s)$ and $(1/sT_{IN}+1)$ at the chosen value of loop gain. An input time constant of 10^{-6} is about the lowest which the amplifier normally encounters, e.g., on the 10^{-8}-ampere range with 100-$\mu\mu$f input capacity, $R_F C_{IN} = 10^{-6}$.

Although the instrument will be unstable with this value of T_{IN} and a gain of 50,000, a reduction in gain to about 20,000 will insure stability. Alternatively, 50 or more feet of cable or a 1,000-$\mu\mu$f capacitor at the input will maintain stability, since T_{IN} will be larger than 10^{-5} seconds.

With an amplifier characteristic of relatively constant phase, it follows that the phase margin is similarly fixed over wide variations of the input time constant. This means that the transient response stays fairly uniform through the entire current range of the instrument. Overshoot to a step input of current on any range is less than 30%. This type of response is adequate for most reactor instrumentation and control applications.

Returning to the effect of T_F in the feedback path, the case where this time constant becomes limiting on the closed loop response is of interest. Significant influence of this lead term is to be expected when it is located about as indicated on the open loop curves of Fig. 6 for representative input time constants. The result, of course, is increased phase margin and damping which would, in general, be selected for damping factors in the vicinity of 0.7: approximately the situation of the $T_{IN}=1$ curve.

A difficulty is immediately evident: note that as R_F is reduced by factors of 10, the instrument tends to become underdamped because T_F reduces by 10 while ω_N increases by only a factor of 3.72. Thus, a fixed value of C_F will not give optimum response on every range, but rather a separate fixed capacitor must be used for each range setting. Even this is not completely satisfactory because the response on a single range will change with input capacity.

Regardless of any need for it, damping will exist on the 10^{-11}-ampere range because a 10^{12}-ohm resistor is used in order that the output span be 10 volts. Ordinarily, insignificant circuit and resistor capacities of less than 1.0 $\mu\mu$f are now important in shunt across the feedback resistor. By special construction these have been reduced to 0.06 $\mu\mu$f so that the instrument time constant with minimum input capacity is about 60 milliseconds. With 2,000 $\mu\mu$f on the

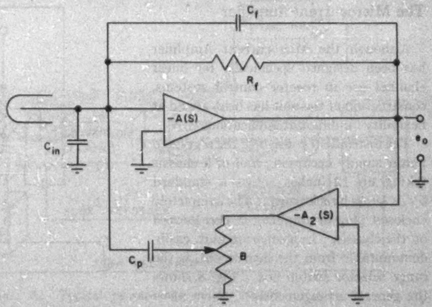


Fig. 7. System schematic for capacity cancellation with positive feedback

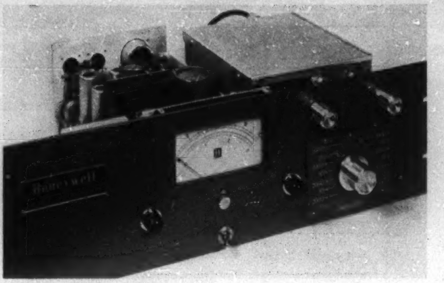


Fig. 8. Experimental model of the Microcurrent Amplifier

input the instrument time constant is less than 150 milliseconds.

For some laboratory applications wide bandwidth at lower current levels than 10^{-11} amperes full scale might be desirable.

If a 10^{13}-ohm feedback resistor must be used, the time constant in the feedback, even with only 0.06 $\mu\mu$f capacity, would be 600 milliseconds. Fig. 7 shows a scheme whereby the dominant lag in the feedback at these low levels might be cancelled out with positive feedback. The approximate condition for cancellation is

$$C_p = \frac{C_f}{A_2\beta}$$

If the positive feedback is made greater than that necessary to cancel the feedback time constant, the stability of the circuit will be decreased and the result is severe ringing or oscillation. This method is, therefore, only applicable for use with this amplifier on ranges lower lower than 10^{-11} because T_F is not, in general, the dominant term and because of the care that must be used in setting the amount of positive feedback as ranges are changed. Variations of C_F, C_P, A_2 and β set a limit to the degree of compensation over a period of time.

The Microcurrent Amplifier

Although the Microcurrent Amplifier has been designed specifically for linear channel use in reactor control systems, construction of the unit has been aimed at flexibility for different applications.

The instrument consists of the regulated power supply occupying half of a chassis taking up 12 inches across a standard $5^1/_4\times19$-inch rack panel. The completely enclosed plug-in amplifier is also located on the chassis. In a separate unit, easily demountable from the instrument, is the range selector switch box. Fig. 8 shows the general arrangement.

Amplifier

The amplifier proper is composed of four direct-coupled stages as shown in the circuit diagram of Fig. 9. Failure to demonstrate significant advantages in drift reduction by differential stages induced selection of a simple single-ended input stage configuration. Plate, screen, bias, and filament potentials are all doubly regulated for stability. The drift that occurs is largely that of electrometer tube contact potential which is not easily compensated or balanced due to variation from one tube to the next.

Two to 5 millivolts per day drift is common, becoming somewhat better with time, for normal variations in room temperature and an unregulated line. The first two stages and their associated components are built on a shock-mounted printed circuit board. Isolation from shock and vibration contributes both to low microphonics and prevention of aging cycles in the electrometer tube. Grid current drops to less than 5×10^{-14} amperes after a settling period of a few hours and open input peak-to-peak noise is below 1%.

A second printed circuit board is used for construction of the third stage of gain and the cathode follower. Output swing is nominally ±10 volts at 5 milliamperes, but can exceed ±20 volts at reduced load.

Forward gain at d-c is 50,000 maximum, but may be varied by means of potentiometer adjustment of third stage cathode degeneration over the range of 5,000 to 50,000. Shaping of the amplifier attenuation characteristic to an average slope of 15 db per decade and phase shift of 67.5 ±10° between 10 cycles and 100 kc (Fig. 3) is accomplished with four phase lag networks placed in the first two stages together with compensation of the interstage dividers. Before insertion of the networks the gain-bandwidth product of each stage was maximized and the interstage capacities reduced as far as possible. The use of a printed circuit was helpful in this regard.

Range Selector Box

Current range is selected by changing R_F, the feedback resistor, in steps of 1, 3, 10, etc. A short-circuiting 18-position 2-deck steatite switch was chosen for its general ruggedness, low leakage, and small strain on insulators during transfer from one position to the next. The last feature and the addition of charge dissipating resistors enables attaining ver low switching transients. Microphonics are substantially reduced by stoutness of construction and tying down most components. Shunt capacity and its damping effects are minimized by sielding switch decks and circuitry wherever feasible. A value of 0.06 $\mu\mu$f can be reached without great difficulty.

In the interest of flexibility a fixed 20-$\mu\mu$f capacitor is added from the summing point to the wiper of a potentiometer between output and ground. This yields a variable damping control for cases where only one or two ranges at a time are of

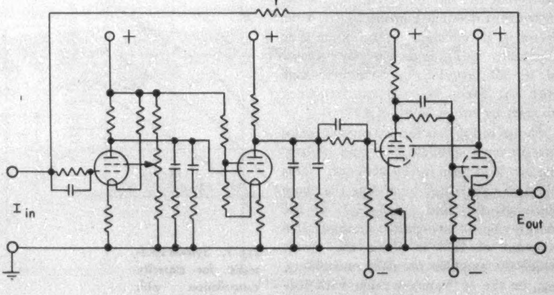


Fig. 9. Circuit of the Microcurrent Amplifier

concern. The entire assembly in its box can be demounted and moved to a remote location, for instance a console, when such is desired for operator convenience.

Performance

A summary of characteristics of the complete instrument is given in Table I. Lines (1) to (3) are for the standard arrangements with a 10-volt feedback span. Choice of a 1-volt span or less yields higher speeds at the expense of stability as shown by (4) and (4A), or lower ranges, as in lines (5), (6), and (6A). All data were taken with maximum gain and a 10-volt output span, i.e., reduced feedback spans were obtained with a divider at the output.

Table I. Characteristics of the Microcurrent Amplifier

Range, Amperes	R_F, Ohms	Feedback Span, Volts	Stability, Per Cent per Day	67% Response, Msec	C_{INB}
(1) 10^{-11}	10^{12}	10	0.05	80	50
(1A) 10^{-11}	10^{12}	10	0.05	150	2,000
(2) 10^{-9}	10^{10}	10	0.05	6	2,000
(3) 10^{-3}	10^{4}	10	0.05	0.025	2,000
(4) 10^{-11}	10^{12}	1	0.5	20	50
(4A) 10^{-11}	10^{11}	1	0.5	100	2,000
(5) 10^{-12}	10^{12}	1	0.5	800	2,000
(6) 10^{-13}	10^{12}	0.1	5	500	50
(6A) 10^{-13}	10^{12}	0.1	5	4000	2,000

Log-N Amplifier

One extension of the Microcurrent Amplifier should be noted. By replacing the range selector switch and resistor assembly with a logarithmic diode, a high-speed log-N amplifier results. Models have been built covering the same 10^{-11}-to 10^{-3} ampere current range with an accuracy of 2 % of full scale. Physically, the substitute circuits—diode, calibrating adjustments, etc.—are contained in a box fitting the same space as the range selector. A period meter and trip circuits are generally combined to furnish a complete intermediate channel.

Magnetic Automatic Power-Range Control for an Aircraft Nuclear Reactor

J. A. RUSSELL
NONMEMBER AIEE

S. F. HEMMENWAY
ASSOCIATE MEMBER AIEE

J. L. SCHARF
ASSOCIATE MEMBER AIEE

P. C. SHARR
NONMEMBER AIEE

Synopsis: Control system philosophy for an aircraft nuclear power plant is discussed briefly. Advantages are presented for a power-range magnetic control system with a minimum number of moving parts. Control system elements and computing devices are discussed in detail. Performance of a breadboard version of the control system is evaluated.

THE POWER-RANGE reactor control system in a nuclear-powered airplane controls the usable power output of the power plant. In the nuclear turbojet engine being developed by the General Electric Aircraft Nuclear Propulsion Department, the reactor takes the place of the chemical-fuel combustor used in conventional turbojet engines. The reactor must supply the correct amount of heat, or power output, to satisfy the over-all engine demands. Therefore, temperature is the primary controlled variable upon which power plant control is based.

A secondary control loop based on reactor flux is integrated with the temperature control loop. If reactor reactivity were controlled by temperature alone, unstable operation could result. The unstable operation could result from an excessive time lag between the actual generation of additional heat in the fuel elements and the electrical signal from temperature sensors in the effluent. Fast response is desirable, and accurate control is essential to assure maximum utilization of the reactor.

A tertiary loop, a reactivity-controlling position loop, is driven from the flux-loop error signal.

The control system is shown in the block diagram of Fig. 1, in which some of the items require explanation. An integration appears between the flux-loop and temperature-loop error to bring the steady-state temperature error to essentially zero. The flux loop is logarithmic to maintain system gain relatively constant regardless of reactor "gain." One of the safety functions is shown, a device which prevents the operator from calling for withdrawal of the reactivity controller under certain conditions. A constant current reference supplies bias, demands, and reference current for the entire system.

Constant current, rather than constant voltage, is required for a magnetic-amplifier system since the control windings are

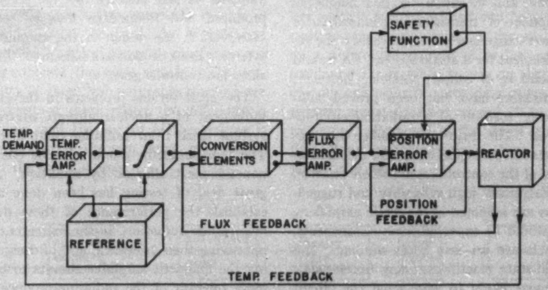
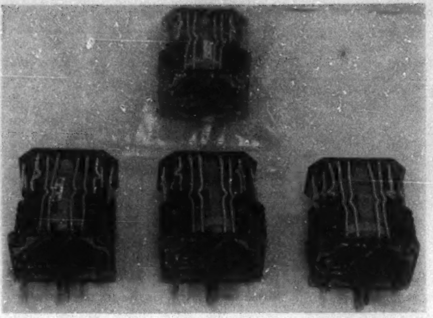


Fig. 1. Typical power-range reactor control

Paper 60-515, recommended by the AIEE Nucleonic and Radiation Instrumentation Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Nuclear Engineering and Science Conference, Cleveland, Ohio, April 6–9, 1959. Manuscript submitted August 4, 1959; made available for printing February 9, 1960.

J. A. RUSSELL, S. F. HEMMENWAY, J. L. SCHARF, and P. C. SHARR are with the General Electric Company, Cincinnati, Ohio.



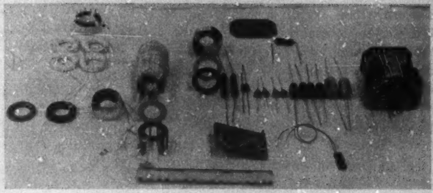


Fig. 2 (left). Unencapsulated amplifiers

Fig. 3 (above). Exploded view of amplifier

sensitive to ampere-turns. Ambient temperature variations in a constant-voltage system would not result in constant ampere-turns; thus reference points would be coupled directly to ambient temperature.

Implementation of Control System With Magnetic Amplifier

Choice of Magnetic Amplifier for the System

Magnetic amplifiers were chosen for this control system because of their extreme reliability. As stated, the power-range reactor control system is a vital part of the total control system for a nuclear aircraft engine. The requirements of aircraft reactor controls are necessarily different from those for non-air-borne reactors. The use of magnetic amplifiers can contribute to the reliability of the system.

When reliability is the foremost design consideration, a so-called "static" system results; i.e., a minimum of moving parts, sliding contacts, and switching contacts is used. A control system in a nuclear engine will be subjected to conditions that include radiation, high temperatures, high acoustic noise levels, shock, and vibration. Either magnetic-amplifier or transistor mechanization of power-range control could satisfy the requirement for a static system. A hybrid system is, of course, possible; however, transistors have not been proved sufficiently tolerant of a radiation environment. The magnetic amplifier thus appears to be a logical choice for mechanization of the control system, where feasible.

Extremely high reliability and ruggedness are inherent in magnetic amplifiers, as would be expected from its elements, which are an iron core, windings, and solid-state reactifiers. Any unreliability is more likely to arise from the circuit design, construction methods, and materials. These considerations are also the means by which the inherent high reliability of these devices can be realized, and a great deal of effort is being directed to this end.

In addition to the reliability aspects of the device, several other salient features substantiate its use for the power-range control system. The size and weight of a magnetic-amplifier system can be equal to or less than an equivalent electron tube implementation since no d-c power supply is needed. Extremely high power gains can be realized in one stage. This gain can be proportioned between voltage and current gain by proper choice of turns to meet the requirements. Many new circuit possibilities are presented by saturable core devices, such as counters, integrators, static switches, and storage elements, to name a few.

As for environment, the magnetic amplifier is quite tolerant of conditions that will be encountered in a nuclear aircraft. Vibration will not damage an encapsulated magnetic amplifier with suitable core cushioning. A newly encountered problem with large turbojets is high-level acoustic noise. Because of its inherent ruggedness the magnetic amplifier is not affected by the forces produced and frequencies encountered. However, in the design of the amplifier attention must be paid to a solid mounting and a low center of gravity.

The most serious problems in the environment of a nuclear aircraft control system arise from radiation and heat. The semiconductor diodes are the components most affected by radiation. A great deal of testing has been done to establish the performance of these devices under radiation, in the interests of improving their tolerance, and of designing the magnetic-amplifier circuits to be more tolerant of the radiation-produced diode changes. Component parts used in these magnetic circuits are capable of operation in ambient temperatures of 125 degrees centigrade.

Encapsulation and wire-combinations have been tested under radiation and temperature. Complete amplifiers have been irradiated to estimated aircraft doses, and changes have not been significant. Circuit developments have recently increased tolerance considerably beyond this radiation level. In view of such testing, the magnetic amplifier appears to be well suited to the specialized environmental conditions encountered in a nuclear aircraft engine control system. Unencapsulated units incorporating the latest circuit developments are shown in Fig. 2, and an exploded view of the same type of device in Fig. 3.

Growth potential of magnetic devices is good, considering the more rigorous radiation and temperature environments of the future. Advanced development of magnetic circuitry without the use of semiconductors is being carried on, as well as significant improvements in the semiconductors themselves. New developments in high-temperature semiconductors will also extend the capabilities of magnetic devices greatly.

Advantages of Magnetic Amplifiers

In general, the magnetic system can be packaged in half the volume of a corresponding electron-tube system with power supplies. The weight comparison shows the magnetic circuitry to weigh in the order of one third the weight of a comparable tube system.

An electron-tube circuit which is used in a complex control system usually requires three or more extrenal sources of power, an a-c filament supply, and two or more high-voltage d-c supplies. The magnetic amplifier is powered by one alternating voltage source (in the case of military aircraft, 115 volts, 400 cycles). There is no need for voltage or a-c to d-c conversion, with the exception of certain center-tapped circuits that require a small integral transformer.

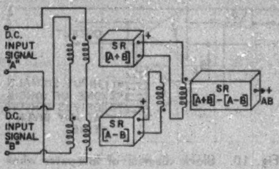


Fig. 4. Block diagram of magnetic relay

To make a comparison of the various power losses and efficiencies between a tube circuit and a magnetic amplifier, it is desirable to look at a particular circuit requirement where power losses become significant. In the control system under discussion, the torque motor valve amplifier, which delivers 3 to 5 watts, is an example.

This particular design has power losses in the copper windings, a bias resistor, and a resistor in series with gate windings. The efficiency of the device, defined as the ratio of power to the load to power taken from the line, is about 70%.

An equivalent d-c tube amplifier delivering 3 to 5 watts and operating class A or class B can be expected to operate at 30% to 50% plate efficiency. When one considers the total power taken from the line, including filament power losses in the power supply and filters, the over-all power efficiency of the amplifier and its associated circuitry is quite small, perhaps 5% to 20%.

The question of optimization arises if heat generated is a problem. A magnetic circuit for high power requirements can have high efficiency. However, a tube circuit, by its physical nature, can obtain a moderately low efficiency.

UNUSUAL CIRCUIT FEATURES

There are some unusual circuit features of magnetic devices that are put to good use in this system. First, it is possible to have multiple electrically isolated inputs to an amplifier. This arrangement is used in the flux amplifier, the temperature amplifier, and the feedback amplifier around the integrator. Each of these amplifiers is at a summing point in the circuit, and through the use of these multiple inputs, the summing can be done without danger of interaction between the various signal circuits. Another feature of magnetic devices closely related to this one is that it is possible to have feedback around an amplifier and still have the input isolated electrically from the output. This method of feedback was used in many places throughout the system.

Magnetic amplifiers also can have very low input resistance, which makes them excellent thermocouple amplifiers because good impedance matching can be had. This property is used in the temperature loop.

Another useful feature of one type of magnetic amplifier, a half-wave amplifier, is that it can sum a-c and d-c signals at the same time. Thus, a-c position feedback around the position loop amplifiers is summed with the d-c demand, and the necessity of demodulating the a-c transducer signal is eliminated (an a-c transducer is the most reliable position sensor).

Particular Circuits and Devices

COMPUTING FUNCTIONS

Continuously Variable Comparative Limiter

The circuit under discussion selects the smallest signal from two input signals, and doubles it. The logic of the circuit is based on the mathematical fact that if the absolute value of the difference between two quantities is subtracted from the sum of the same two quantities, the result is always twice the smaller: where B is the smaller

$$|A+B|-|A-B|=2B$$

Because of the absolute value signs, if A is smaller, the result is

$$|A+B|-|B-A|=2A$$

This logic is implemented by the circuitry shown in the block diagram of Fig. 4. Saturable reactors are ideal for the summing and subtracting since their output is always of one polarity and hence represents the absolute value of the sum or difference. One saturable reactor is used to sum the two input signals and another identical saturable reactor is used to subtract the same signals. The outputs of these two saturable reactors are fed into a third, where the difference is taken, and the output of this reactor is proportional through some gain factor to the lower signals.

It is easily seen that this circuit makes an ideal limiter for either a fixed limit or a continuously variable limit. If one input signal is set at some fixed level, the output of the circuit will follow the variable signal from zero up to the level of the fixed signal. There the output will cease to follow the increasing variable signal and instead stay at the level dictated by the fixed signal. This action is shown in Fig. 5. Since the output of the circuit always follows the lower of the input signals, some variable limit signal can be fed into one input and thus cause the maximum output to stay at a level below or equal to that dictated by the limit signal.

Signal Comparison

It is necessary in the control system to perform many switching operations where one signal A, being larger than another B, will cause the switching to occur, but signal B, being larger than A, will not. Furthermore, in many cases it is desired that the differential signal that causes switching be very small, in the μamp (microampere) region, if the signals are currents.

The magnetic relay circuit was developed to perform just such switching operations; the circuit diagram is shown in Fig. 6. Briefly, the circuit operates as follows: The magnetic amplifier raises the input signal from the μamp region into the milliamp region. A diode gives unipolarity switching. The saturable reactor a-c output drives a ferroresonant trip circuit. The magnitude of this alternating voltage is controlled by the direct current in the saturable reactor control winding. The ferroresonant circuit is a bistable device that trips at a certain value of alternating voltage. When the circuit is tripped, its impedance becomes very low and a large current is allowed to flow through it and the relay coil in series.

This circuit has a trip point that is independent of the relay trip point and line voltage frequency, and can be compensated for temperature. It can switch on a differential current of as low as 1 μamp and can be made to trip on signals of only one polarity (providing opposite polarity signal does not exceed 20 to 25 milliamp). Removal of the series rectifier allows switching action from any polarity of signal.

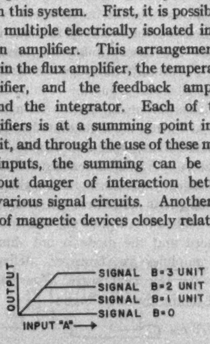


Fig. 5. Characteristic of variable limiting

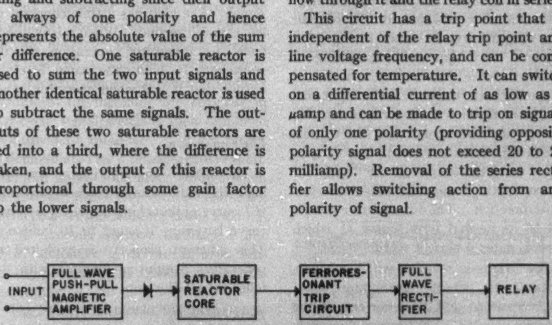

INPUT
FULL WAVE PUSH-PULL MAGNETIC AMPLIFIER
SATURABLE REACTOR CORE
FERRORESONANT TRIP CIRCUIT
FULL WAVE RECTIFIER
RELAY

Fig. 6. Continuously variable comparative limiter

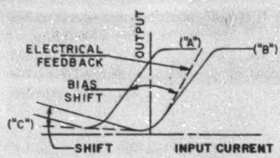


Fig. 7. Alteration of basic amplifier characteristics to act as magnetic rectifier

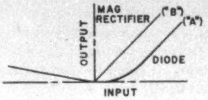


Fig. 8. Comparison of semiconductor rectifier and magnetic rectifier

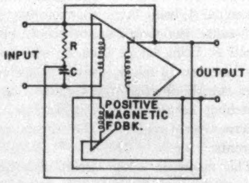


Fig. 9. Functional diagram of integrator

Current Blocking

A device is needed in the control system that blocks currents of one polarity while passing those of opposite polarity. This device must have a current input versus current output characteristic that is linear from zero input to some maximum input. This requirement eliminates the use of a diode because at least 0.4 to 0.5 volt across it is required to open it.

The magnetic rectifier was developed to meet the need for such a device. This circuit consists of a square circuit magnetic amplifier modified for d-c output. The output curve is shown in Fig. 7 as curve *A*. By appropriate shunting of diodes, the knee of the curve can be shifted to the zero axis, curve *B*. The input current is fed through a resistor in series with the output. The polarity of the voltage across this resistor is such that it subtracts from the output in the "back" or nonconducting region and adds in the forward direction. The output curve is rotated to its desired form, curve *C*, which approximates a perfect rectifier.

The magnetic rectifier will block currents of one polarity (maximum "leakage," 5 μamp) and pass current of the opposite polarity. The forward characteristics are quite linear from 0.0 to 0.8 milliamp. Fig. 8 compares a diode, curve *A*, with the magnetic rectifier, curve *B*.

Integrator

The control system requires an integrator with a linear integrating action, with respect to time, that will hold at the particular level reached if the input signal is suddenly removed. This integrator must also be capable of acting as a straight amplifier under certain control conditions.

The circuit meeting these requirements is shown in Fig. 9. A full-wave center-tap magnetic amplifier is used as the computing amplifier and has considerable positive feedback to provide essentially infinite gain. The integrating action is controlled by the resistance-capacitance combination across the input, the product *RC* determining the rate of integration for a particular input. If the signal is removed, the circuit provides holding action, since capacitor charge leaks through the amplifier windings in a direction that causes the amplifier output to oppose capacitor discharge. The output is taken from the amplifier terminals.

This circuit arrangement also has the advantage that very little current flows in the input winding of the amplifier because of the electrical feedback of the output. The circuit will "servo" in to the point where the current into the input winding is essentially zero. The output is also forced by this arrangement to follow the charging of the capacitor, giving quite linear integrating action over a certain range.

To cause the integrator to perform as an amplifier, the circuit of Fig. 10 is used. It is a simple servo loop in which the amplifier *K* detects the error between the integrator output and the auxiliary input, amplifies it, and feeds it back into the input of the integrator. Thus, the integrator output, operating as an amplifier, is constrained to follow the input signal. The loop has a time constant of less than 0.1 second. When it is switched back to integrator operation, it instantaneously holds the same level of output that it had when operating as an amplifier.

Log Conversion

The forward characteristic of an ordinary silicon diode has a nearly logarithmic *E-I* curve approaching a true logarithmic curve between 10 μamp to 10 milliamp. This inherent property is exploited to generate a computing function with the following relationship:

$E = K_1 \ln i + K_2$

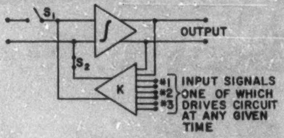


Fig. 10. Block diagram of integrator when switched to act as an amplifier

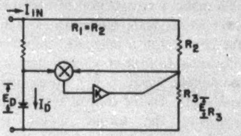


Fig. 11. Log conversion bridge

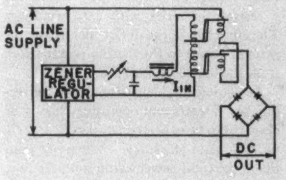


Fig. 12. Constant current reference

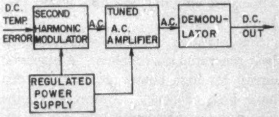


Fig. 13. Temperature amplifier

The magnetic amplifier following the diode must not disturb the *E-I* relationship of the diode by loading and must furnish power to the next stage.

Because of the low level of diode input current and susceptibility to noise, the diode is operated in a bridge circuit, shown in Fig. 11.

When current I_D flows through the diode, a voltage E_D is generated. This is compared to the magnetic-amplifier output voltage E_{R3} by a sensing coil across the bridge. The sensing coil is the input to a magnetic amplifier with very high gain. When E_{R3} equals E_D, the bridge is balanced and the diode is not shunted. The amplifier is, however, delivering power to the next stage at a voltage equal to the diode voltage:

$I_D = 1/2\ I_{in}$

Since the input current to the bridge

Vol.
Pt.
196

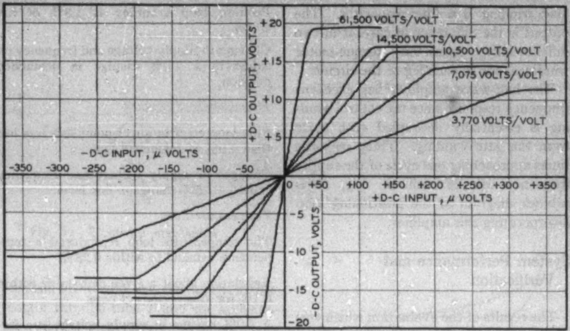


Fig. 14. Second-harmonic modulator electronic d-c amplifier. Performance curve showing various amounts of negative feedback

splits between R_1 and R_2, the voltage-current relationship now becomes

$$E_{R3}=E_D=K_1 \ln 1/2\, I_D+K_2$$

But

$$\ln 1/2\, I_D=\ln I_D+\ln 1/2$$

Now, let

$$K_3=K_2+\ln 1/2$$

Then

$$E_{R3}=E_D=K_1 \ln I_D+K_3$$

This relationship holds for 2.5 to 3 decades of input current.

OTHER MAGNETIC DEVICES

Constant Current Reference

The basic element of the constant current reference is the saturable reactor with a bridge rectifier in its output. It has the property of proving a direct current independent of the load resistance and small changes in line voltage and frequency; it is controlled only by the input current.

The transformed current ratio is an inverse function of turns ratio only. Current gains of the order to 5 to 20 are very easily obtained.

The control winding of the saturable reactor is driven by a regulated small current supply, such as the output of zener diode. The total components form an open loop and furnish a large current output independent of changes in line voltage supply, line frequency, and load resistance limited by

$$I_0\, R_L=1/2\, E_{\text{gate}}$$

By optimizing each unit, current regulation of 0.005% can be obtained; see Fig. 12.

Temperature Error Amplifier

The difference of temperature demand and feedback is amplified by the temperature error amplifier. The zero and gain stability of this amplifier is such that it could be used directly in the temperature feedback line. However, if this is done, the variations in gain and zero of the amplifier cannot be differentiated from the true feedback signal. This would result in changes in the controlled variable to compensate for these variations. An error amplifier causes only a change in accuracy of the loop by a change in gain. Zero drift in the amplifier will again cause error in the controlled variable to compensate for it, but less error than if it were in the feedback. A more stable and accurate loop will thus result from using the amplifier as an error amplifier.

This particular design more than adequately meets requirements by employing magnetic and electronic devices. Fig. 13 is a block diagram of the amplifier.

A second-harmonic magnetic modulator modulates the incoming d-c error to an 800-cycle signal which is then amplified in a tuned a-c amplifier. The a-c output is demodulated by a phase-sensitive rectifier, and appears as direct voltage again in the output. The d-c is fed back through a second winding in the modulator as over-all negative feedback. The entire amplifier is driven from a self-contained saturable-transformer-regulated supply.

The second-harmonic modulator provides considerable power gain, besides performing the modulating function, and in addition it has extremely low null drift. The theoretical minimum detectable signal of a magnetic second-harmonic modulator approaches 10^{-13} watt. About 10^{-14} watt has been achieved in this design. The a-c amplification, of course, does not add any null drift. Very small errors can be sensed consistently to provide excellent loop sensitivity with little drift. Loop accuracy with this amplifier is held to better than 0.25% of the controlled variable. A side feature, which is unnecessary in an error amplifier but useful in other applications, is the linearity of the amplifier with and without feedback. This is shown in Fig. 14, where several values of feedback have been set. The gain may be set from 61,000 volts per volt to 3,800 volts per volt. It is of interest to note that when magnetic feedback is used, there is no galvanic coupling between the output and input, resulting in freedom from grounding and isolation difficulties. It has been found that the modulator is capable of accepting an overload of greater than 100 times its normal maximum input with no permanent effects.

The open-loop gain does not vary more than 1.0% or 2.0% from 107 to 125 a-c input. Response time with nominal negative feedback applied is approximately 15 milliseconds. The drift of the entire circuit over a period of 24 hours was measured as approximately 2 μvolts referred to the input.

Voltage Current Conversions

When a 2-input-winding full-wave magnetic amplifier is connected as shown in Fig. 15, a closed-loop servo is formed. This arrangement provides an output affected only by the varying input voltage signal. Hence

$$N_1\, I_1=N_2\, I_2$$

independent of R_2, provided I_2R_2 does not exceed one fourth gate voltage. By decreasing N_2 the output current must increase to maintain balance and therefore a current gain can be obtained. An applied voltage at the input terminals can thus generate a constant current in the load, independent of load resistance (within limits, as already mentioned) but

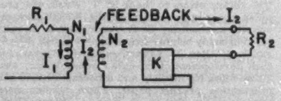


Fig. 15. Voltage current conversion circuit

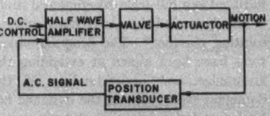


Fig. 16. Position loop block diagram

directly proportional to the applied voltage signal.

Valve Torque-Motor Drive

The torque-motor valve amplifier is a half-wave amplifier used in a closed-position loop. The device has two input windings. The input or demand is a d-c position demand signal. The second winding accepts an a-c signal from a linear differential transformer feedback, a function of output position. This circuit compares an a-c signal to a d-c signal without the use of demodulators. A half-wave amplifier, except for a difference in gain, will operate an a-c or d-c input.

The circuit itself consists of two cores arranged in a half-wave half-bridge. Each core has a gate winding and a reset winding. Common to both cores are the d-c and a-c control windings and a negative feedback winding. Across each load winding is a filter capacitor. The output is the difference in current in two differentially connected torque-motor windings, one in each leg of the circuit.

The half-wave amplifier has excellent frequency response since the control winding is essentially decoupled each cycle from the gate winding. Thus, response times approaching one cycle of the supply frequency can be realized. Fig. 16 is a block diagram of the positioning loop incorporating this amplifier.

System Performance and Verification

The results of the evaluation which was conducted on the magnetic-amplifier power-range control system are summarized below.

Position Loop

Position loop noise is 0.1% of full travel.

Position loop accuracy is 1.3% of full travel.

Changes in supply voltage and frequency resulted in a 1.0% change in demanded position.

Flux Loop

Flux loop error in an 11-point test was less than 1.0% of demanded flux.

Changes in supply voltage and frequency caused a 1.5% change in flux level.

Temperature Loop

The temperature loop reproduced a temperature demand to within 0.28%.

Regulation about a given demand is within 0.1% for long periods of time.

A 5.0% change in supply voltage and frequency caused an error of 1.4% of demanded temperature.

The control system fulfilled the transient requirements imposed on it by the desired reactor dynamics.

Instrumentation for Experimental Reactor Kinetic Studies

EDWARD L. GARDNER
NONMEMBER AIEE

REACTOR POWER excursions involving stable periods ranging from a few seconds to 2 ms (milliseconds) have been experimentally investigated in the course of the Kinetic Experiments on Water Boilers (KEWB) program conducted by Atomics International for the U.S. Atomic Energy Commission. Reactivity introduction, power and temperature measurement, and data analysis have imposed some unique requirements in instrumentation and special apparatus.

Early experimental investigations of the dynamic behavior of the water boiler type of reactors were conducted on the SUPO reactor at Los Alamos by Kasten,[1] and on the Livermore reactor by Flora, and others.[2] These experiments included reactivity relase up to a maximum of only about 0.4%, resulting in minimum reactor periods of about 6 seconds.

The present efforts on KEWB were initiated as a full-scale experimental study of the dynamic behavior of homogeneous solution reactors.[3] The program objectives have been aimed at extending the knowledge and understanding of the dominant parameters in the dynamic behavior and the safety of these types of reactors. The operating parameters which bear directly on the transient behavior of the reactor and which are under investigation include the amount of reactivity released, the initial fuel temperature, the initial core pressure, and the initial power level.

The approach in the program has been to construct a prototype reactor with the special provisions required for the experimental transient studies. This prototype reactor is installed in a special test facility, which includes an underground building to house the reactor test cores with reflector, gas-handling system and associated equipment, a waste disposal system, a small building containing amplifiers and power supplies, and the control building, some 200 feet from the test building. A comprehensive description of the facility is contained in reference 4. The instrumentation, control, and data treatment aspects of this program which are considered unique are discussed in this paper.

Reactivity Mechanism

Two means are used to insert reactivity into the KEWB reactor. The safety and control rods are withdrawn singly or in groups of two, three, or four. The low rate of reactivity insertion by this means limits its usefulness to effective step inputs of up to 0.7% or to ramp reactivity studies involving insertion rates of 0.12%/sec (per second). Reactivity releases approaching 4% produce reactor periods down to 2 ms, so that it is necessary to insert reactivity as nearly approaching a step input as possible. To achieve this rapid reactivity insertion in the KEWB a pneumatic accelerator is used to eject a poison rod from the core. This device can insert 4% reactivity in 20 ms. This near-step reactivity release places the reactor on a stable period of 2 ms and could be satisfactorily used to produce even shorter periods.

The pneumatic accelerator, Fig. 1, is comprised of a poison rod P attached to an air-driven piston p, a compressed air reservoir R provided with a solenoid S operated valve V_0, and a cylinder C fitted with various valves. The poison

---

Paper 60-514, recommended by the AIEE Nucleonic and Radiation Instrumentation Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Nuclear Engineering and Science Conference, Cleveland, Ohio, April 6-9, 1959. Manuscript submitted August 4, 1959; made available for printing February 9, 1960.

EDWARD L. GARDNER is with Atomic International Division of North American Aviation, Inc., Canoga Park, Calif.

The author wishes to thank M. T. Webb, who provided the original design of the KEUB fast reactivity mechanism; D. P. Gamble and R. K. Stitt for their many helpful suggestions for circuit design; and R. J. Allen, D. Vawter, and R. B. Wimmer for their development of the data readout device; also J. W. Flora for his help and encouragement in the preparation of this paper. These studies were conducted for the U. S. Atomic Energy Commission under Contract *AT(11-1)-GEN-8*.

Vol.
Pt. I
196

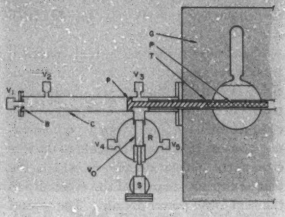


Fig. 1. Reactivity mechanism

rod and its piston are contained in the reactor through-tube T and the accelerator cylinder respectively. The through-tube, and therefore the poison rod, pass through the core on a horizontal diameter.

The piston is driven by compressed air released into the cylinder from the reservoir. It draws the poison rod completely out of the reactor reflector G, then stops the rod on an air buffer.

As the piston is driven outward by the compressed air from the reservoir, it compresses the air in the cylinder ahead of it. This air compression acts as a buffer to slow down and stop the piston. The compressed buffer air would also drive the piston and the attached poison rod inward in such a manner that the piston and rod would oscillate on an air spring if some means were not used to prevent it. The spring-loaded valve V_2 bleeds off the buffer air, allowing the piston to travel outward with the least possible deceleration while the poison rod is inside the reflector. The size of the valve orifice determines in large measure the rod travel characteristics. If the orifice is too large the piston will strike the end of the cylinder. On the other hand, too small an orifice will cause the rod to rebound into the reactor. The appropriate orifice size was determined with the aid of Fastax camera films of trial ejections and was selected to give the least ejection time compatible with a rebound limited to 6 inches, which does not allow the poison rod to re-enter the reflector. The pressure of the drive air, although reduced, overcomes the rebound and again drives the piston outward and holds it against a rubber bumper B, which prevents the piston from striking the metal end of the the cylinder.

The poison rod is reinserted into the reactor by introducing compressed air at 8 psi (pounds per square inch) into the cylinder on the buffer side of the piston by means of a valve V_1. The spring-loaded dump valve remains closed against this pressure and the piston drives the poison rod gently into the reactor until it reaches a remotely adjustable mechanical stop. The air trapped between the piston and valve V_0 during the reinsertion operation is released through valve V_3, which opens simultaneously with V_1. The position history of the poison rod during an ejection is provided by a magnet carried by the piston. The magnet generates a voltage as it passes under a series of pick-up coils located at 2-inch intervals along the cylinder.

A panel located in the control room console provides for the complete remote control and supervision of the reactivity mechanism.

The piston can withdraw the poison rod from its farthest inserted position to where the rod clears the reflector (30 inches) in 60 ms. The rod withdrawal time and maximum velocity are varied by adjusting the pressure of the air in the reservoir. The rod position traces in Figs. 2 and 3 show the maximum rod velocities to be 800 and 600 inches/sec respectively, which were produced by reservoir air pressures of 80 and 60 psi respectively.

It is necessary from the experimental standpoint to establish the stable reactor periods at as low a reactor power as possible and, in cases involving effective step changes in reactivity, the stable period must prevail for at least one dec-

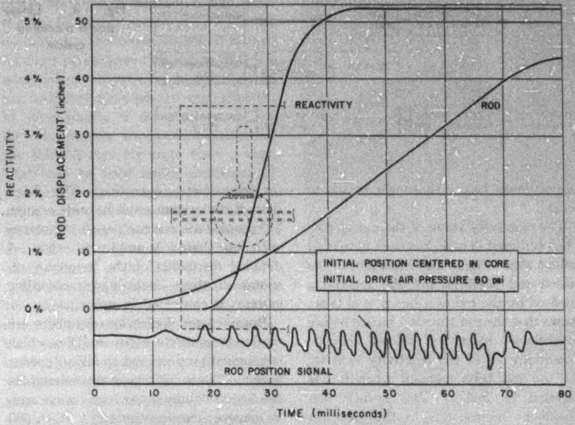


Fig. 2. Poison rod position versus time

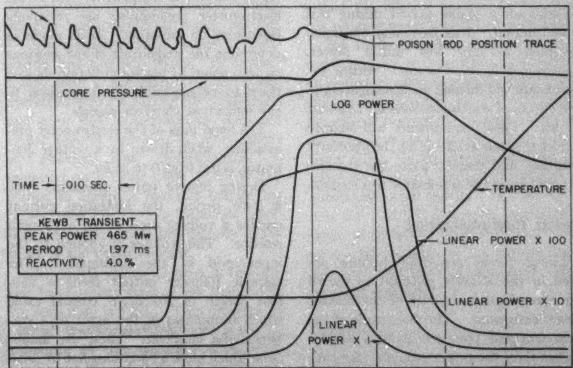


Fig. 3. Power transient oscillogram

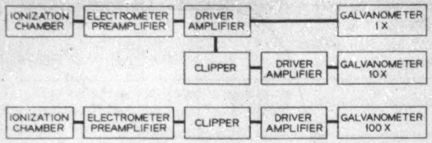


Fig. 4. Linear power recording system

ade of power before shutdown begins to occur.

The reactivity value of the poison rod, when centered in core, has been measured to be 5.2%. Fig. 2, which is a plot of the poison rod position and the reactivity released by the rod as a function of time, shows that the rod travels 4 inches before appreciable reactivity release begins. Essentially all of the reactivity held by the rod has been released when it has traveled 22 inches. The 22-inch displacement corresponds to the eleventh rod position pulse, which is indicated by an arrow in Figs. 2 and 3.

Since the KEWB core is loaded to 4% reactivity, the reactor is not critical until the rod has moved out 8 inches. During this 8-inch travel the rod is accelerated to nearly maximum velocity. This permits the mechanism to insert the 4% reactivity in 20 ms.

Fig. 3 is an oscillogram of a typical reactor power excursion with a 2-ms reactor period. It may be noted that the eleventh pulse, which indicates full reactivity insertion, occurs 43 ms before peak power. An inspection of the more sensitive linear power traces shows that the reactor power was rising with a constant period until 10 ms prior to the power peak. The time increment between these two events, i.e., reactivity fully inserted and onset of shutdown, is 33 ms. The reactor power increased a little over 7 decades on a stable period during this time interval. This particular excursion involved a very low initial power, approximately 10^{-4} watts. Future experiments will include an investigation of the effects of starting powers up to about 25 kw. These experiments will sharply reduce the time available for the reactivity addition, and thereby place the greatest demand upon the accelerator mechanism.

Power Instrumentation

Linear power recording systems are used in the KEWB program to record the three highest decades of the reactor power excursions. The magnitude of the excursion peak powers is about 4 decades higher than the power available for calibration. The usable dynamic ranges of the electrometer should cover 5 decades for convenient operation. The band pass of the system must be wide enough to record 2-ms reactor period excursions with less than 5% amplitude error. A detailed discussion of the frequency response of these linear power recording systems is contained in reference 5.

Boron-coated ionization chambers are used to measure the n (neutron) flux which is assumed proportional to instantaneous power. The sensitivity of these chambers is 8.5×10^{-15} amp/n/sec/cm^2 (where amp = ampere; cm = centimeter). With 600 volts applied to these chambers, 150 μamp can be drawn without saturation effects. Such a current would flow when the flux at the surface of the chamber was 2×10^{10} n/sec/cm^2. Since the flux in the reactor room at the peak of the highest power excursions is 2 or 3 decades greater than 2×10^{10} n/sec/cm^2 the ionization chambers are shielded by cadmium and water-cadmium attenuators.

The electrometer preamplifiers are designed to operate in the reactor room and are connected to the ionization chambers with 4 feet or less of coaxial cable. Fifteen-range resistors in the preamplifier have values starting at 10^8 ohms, and descending to 10^3 ohms. Each decade is split at 5x, 2x, and 1x.

Regulated positive and negative high voltages are supplied from a nearby instrument room through 50 feet of cable. All of the circuit components used in the electrometer preamplifier are relatively insensitive to radiation. Tests made to determine the magnitude of the spurious signals caused by reactor radiation showed them to be negligible in comparison to the ionization chamber signals.

The band pass of the electrometer preamplifier, when driven by a voltage generator, is flat from 0 to 35 kc.

During reactor power transients with a 2-ms period, the ionization current creates a voltage drop across a 10^4-ohm resistor. This voltage is applied to the input grid of the preamplifier. The cathode follower output feeds a 250-foot length of *RG 62/U* coaxial cable which terminates in the input of a stabilized d-c amplifier. These d-c amplifiers, which have a band-pass flat to ±0.3 decibel from zero to 5 kc, are used as driver amplifiers. The galvanometers used have a frequency response flat from zero to 1,200 cycles/sec.

Four linear recording systems are used to measure the reactor power during transient operation. They are grouped in pairs to provide self-monitoring. One channel in each pair drives two galvanometers so that one of the galvanometers covers the power decade in which the transient peak occurs and the other covers the next lower power decade. The second channel of each pair covers the second lower power decade. Block diagrams of the circuits are shown in Fig. 4. In this manner the three upper decades of the power excursion are linearly recorded. It will be noted in Fig. 3 that the two more sensitive linear power signals are clipped to prevent overloading the galvanometer driver amplifiers and to limit the galvanometer currents to safe values.

Temperature Instruments

As energy is released in the water solution of uranyl sulfate, the temperature of the core rises. If the reactor period is short enough, the heat distribution is adiabatic, and until bubble formation causes movement of the solution this temperature distribution in the core is a good indication of the flux distribution. The temperature history during a transient provides data on that part of the intrinsic reactor shutdown due to temperature reactivity compensation. Dynamic measurement of the temperature presents several problems.

The temperature-sensing elements must withstand immersion in the core solution,

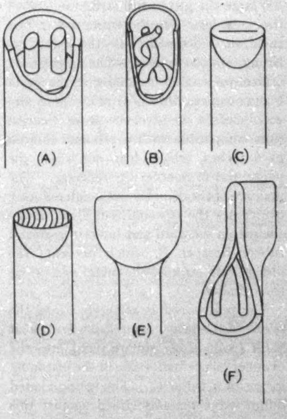


Fig. 5. Thermocouple fabrication sequence

Fig. 6. Digital oscillogram readout system

which is highly corrosive. It is composed of 0.75 molar UO_2SO_4 and 0.5 molar H_2SO_4 in distilled light water. The pH of this solution is 0.7. If the thermocouples are to survive in the core they must be made of corrosion-resistant metals or be protected from direct contact with the fuel solution. The first thermocouples used in the KEWB were 0.005-inch iron and constantan wire insulated with MgO enclosed in a type *347* stainless-steel sheath. The outside diameter of the sheath was 0.040 inch. A suitable length of flawless stock was cut and polished to remove tool marks. The MgO insulation was sandblasted out to a depth of 0.020 inch or more; see Fig. 5(A). The wires were twisted around each other and pressed down 0.005 to 0.010 inch below the end of the sheath, which was closed over the wires with the use of a dull diagonal cutting plier, Fig. 5(C), and sealed by heliarc welding without the use of extraneous metal, Fig. 5(D). The welded tip was polished and inspected for MgO filled pits or other flaws. The good tips were flattened by introducing about 0.060 inch of the sheath between the specially shaped and polished jaws of a vise. This flattening enclosed the thermocouple junction in a blade 0.007 inch thick, 0.1 inch wide, and 0.1 inch long; see Fig. 5(E). After a final polish and inspection, the thermocouples were given electrical tests. Fig. 5(F) shows a section of a completed thermocouple.

Satisfactory thermocouples were then connected to a cold junction and a recording oscillograph. The thermocouple was plunged into a beaker of hot water and the time for the couple to respond to 63% of the final temperature change was measured. This method of obtaining response time is acceptable when dealing with time constants longer than about 5 ms. Time constants of 20 ms were obtained with the best thermocouples produced in this manner.

From the temperature trace of the oscillogram of Fig. 3 it may be seen that the temperature rise in a 2-ms period power excursion is virtually a step function for these thermocouples. While the usefulness of the thermocouples is limited for short-period power transients, they are adequate if the period is longer than 90 ms, and they have proved very effective in other phases of the program. They have resisted the adverse environment of the KEWB core for more than a year. They can be used under conditions of high temperature and pressure as well as as in a corrosive medium. Their response is faster than any commercially available thermocouple which has been offered for use in the KEWB.

A 2-ms period reactor excursion requires that the thermocouples have a time constant of 0.3 ms if the reported temperatures are to be within 10% of the true value at all times. A continuing development program has demonstrated that bare wire thermocouples can be fabricated which have time constants at least as short as 0.5 ms. The useful lifetime of these thermocouples is now being investigated.

Data Reduction

To expedite the detailed reduction of occasional oscillograms under circumstances which are not appropriate for routinely employed digital computer techniques, an oscillogram trace displacement readout device was developed. There are 10 to 15 traces, depending on the experiment, which are recorded on a KEWB oscillogram. These include linear power, log power, temperature, and pressure traces. Ease of operation, accuracy, and digital readout in terms of parameter units were the important design requirements.

The readout device consists of a linear potentiometer, a voltage supply scaling network, and a digital voltmeter; see Fig. 6. The precision linear motion potentiometer has a 5-inch travel. Its repeatability error is less than 1.0% full scale; hysteresis does not exceed 1.0% full scale; and error due to nonlinearity does not exceed plus or minus 0.5%, independent of position. The potentiometer is mounted on a lucite plate which is machined to accept a movable magnifying cursor; see Fig. 7. To eliminate backlash the cursor is attached directly to the slider of the potentiometer.

The digital voltmeter used in this system has linearity error of less than 0.3%. Its resolution is 0.2%. This provides an over-all system precision which is limited primarily by the accuracy with which the cursor is located on the oscillogram trace.

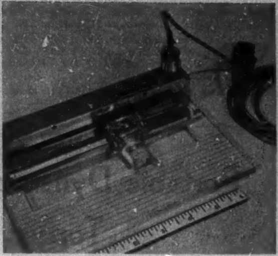


Fig. 7. Readout head showing linear potentiometer and cursor

The output potential may be adjusted to be proportional to power, pressure, temperature, or strain by properly scaling the voltage supplied to the ends of the potentiometer element. The voltage source, which has coarse and fine adjustments, is contained in a compact control unit. The calibration data pertinent to the trace under inspection is used to determine this voltage. The voltage controls have arbitrary scales, and they are adjusted until the desired scale factor, in parameter units per inch, appears on the digital voltmeter. During calibration exactly one fifth of the potentiometer supply voltage, which corresponds to precisely 1-inch trace deflection, is delivered to the voltmeter. The value of any point on the trace is thereafter displayed directly by the voltmeter when the cursor is positioned.

The control unit also provides a bias source which permits the reader to adjust the voltage at any arbitrary displacement of the cursor to zero. This allows the deflection of the trace to be read relative to its own base line without having to adjust the relative position of the readout potentiometer and the oscillogram.

The readout circuit of this device is controlled by a 3-position foot switch. In its first, or normal, position, the calibration factor appears on the voltmeter. Repetitive reading of the scale factor insures against error resulting from instrument drift. In the second position, the value of the trace deflection appears in parameter unit values. In the third position the net deflection potential is made available to auxiliary devices such as a Clary printer or an International Business Machines Corporation card-punching machine.

References

1. REACTOR DYNAMICS OF THE LOS ALAMOS WATER BOILER, Paul R. Kaston. *Chemical Engineering Progress Symposium*, "Nuclear Engineering, Part I,"

American Institute of Chemical Engineers, New York, N. Y., vol. 50, series no. 11, 1954, p. 229.

2. Temperature Effect on Reactivity of the C. R. & D. Water Boiler, J. W. Flora, J. W. Shortall, W. E. Drummond. *LRL-148*, Livermore Research Laboratory, 1954.

3. Program Review of the Water Boiler Reactor Kinetic Experiments, M. E. Remley, J. W. Flora, D. L. Hetrick, L. P. Inglis. *Special Report, NAA-SR-1525*, North American Aviation, Los Angeles, Calif., 1956.

4. Experimental Studies of the Kinetic Behavior of Water Boiler Type Reactors, M. E. Remley, et al. *Proceedings*, Second International Conference on the Peaceful Uses of Atomic Energy, Geneva, Switzerland, 1958.

5. Transfer-Function Response of a High Impedance Nuclear Power Channel, R. J. Allen. *Proceedings*, Nuclear Engineering and Science Conference of the Fifth Nuclear Congress, Cleveland, Ohio, Apr. 5–10, 1959.

Portable Dynamic Electronic Speed-Torque Indicator

HERMAN R. WEED
MEMBER AIEE

Synopsis: The concept of instantaneous observation of the complete speed-torque characteristic of a motor while it accelerates from standstill to full speed is based on the assumption that a large percentage of the torque is utilized in acceleration of the inertia if the motor is otherwise unloaded. The basic instrument needed is an accelerometer with dynamic response capable of indicating the true starting torque and the transient cusps of the torque curve. This paper discusses the application of a 2-phase a-c instrument tachometer and active high-gain operational amplifiers to the design of a portable speed-acceleration instrument. The tachometer utilizes a drag-cup rotor and produces a constant frequency carrier output that has a minimum of ripple with frequency proportional to speed, thus eliminating the problem at low speeds of ripple amplification by differentiation encountered when using conventional tachometers. The constant frequency carrier output simplifies the problem of amplification and noise found with the homopolar-type generator. The use of stabilized operational amplifiers for differentiation and amplification after demodulation both provide the necessary dynamic response for the torque signal and virtually eliminates the problem of drift in the d-c stages. Operational applications of the instrument including torque versus speed, acceleration versus time, and speed versus time are included from the engineering laboratories of Robbins and Myers, Inc., Springfield, Ohio.

THE CONCEPT of instantaneous observation of the complete speed and torque characteristics of a motor while it accelerates from standstill to full speed has been the subject of papers by many competent authors.[1,2] The technique might be summarized as the accurate determination of the instantaneous speed during acceleration, the differentiation of this speed signal, and the assumption that during the acceleration period the major portion of the developed motor torque is used to accelerate the inertia. Thus, the acceleration term is considered proportional to torque. The speed and acceleration signals may then be displayed on the vertical and horizontal axes of an oscilloscope to plot the conventional speed-torque characteristic. The system is thus an accelerometer in basic operation.

Speed and Differentiation Problems

Although simple in concept, the previously stated principles are somewhat more difficult to instrument. Two problems overshadow the others. The first is that of speed measurement. Since only a fraction of a second may be required for the entire acceleration process, the tachometer must be of such nature as to provide an accurate instantaneous indication of speed during this time. This suggests very low inertia and very short electrical time constants. This is particularly true if the instrument is to be portable and not permanently attached to the motor shaft, thus permitting it to function properly other than in the laboratory. The second problem also affects the first. Since the speed signal is to be differentiated to produce the acceleration term, ripple voltages on the speed signal will be emphasized and may easily overshadow the acceleration term itself. This ripple voltage term is difficult to overcome if it has a frequency proportional to speed. In order that the system be capable of indicating cusps in the torque curve, it would be necessary for any filters used to have little attenuation over most of the speed range; thus the emphasized ripple term might be comparable to the torque term. Various types of tachometers have been utilized, such as photosensitive devices measuring the

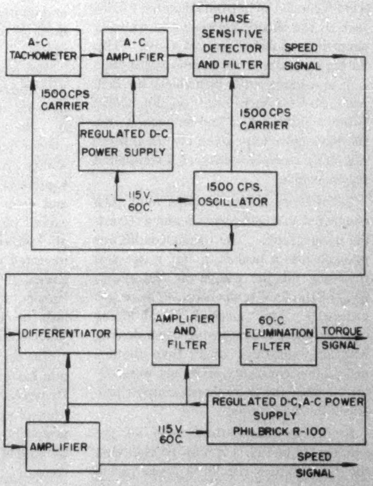


Fig. 1. Block diagram of speed-torque tester

Paper 60-660, recommended by the AIEE Special Instruments and Auxiliary Apparatus Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE North Eastern District Meeting, Providence, R. I., May 2–4, 1960. Manuscript submitted January 7, 1960; made available for printing March 18, 1960.

Herman R. Weed is with Ohio State University, Columbus, Ohio, and Robbins and Meyers, Inc., Springfield, Ohio.

The author wishes to acknowledge the co-operation of Robbins and Meyers, Inc., in making data and laboratory results available for publication. Particular appreciation is due T. C. Lloyd for his enthusiasm and interest in the development of engineering laboratory facilities.

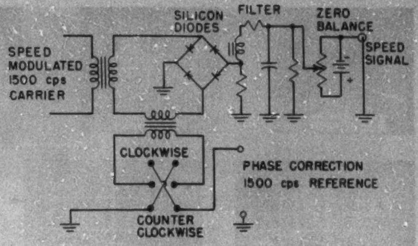

SILICON DIODES
FILTER
ZERO BALANCE
SPEED SIGNAL
SPEED MODULATED 1500 cps CARRIER
CLOCKWISE
PHASE CORRECTION 1500 cps REFERENCE
COUNTER CLOCKWISE

Fig. 2 (left). Speed-torque indicator in operation

Fig. 3 (above). Simplified circuit phase sensitive detector

number of light pulses, magnetic devices measuring voltage pulses, homopolar generators having no commutation ripple but considerable noise, and conventional a-c and d-c tachometers, always with the two problems of low inertia and ripple voltage in mind.

The second problem of differentiation also presents the question of how the differentiation will be accomplished. Until recently, sufficiently stable d-c amplifiers were not available to permit other than approximate differentiation with passive circuit components and the associated limitations in initial values and frequency response. If a torque signal was desired down to zero speed, d-c amplifiers of some design were still required with their associated problems of drift, calibration, and stability.

More or less successful solutions to these problems have been devised and some of the authors previously mentioned discuss workable systems. It is the purpose of this paper to present a solution that more adequately meets all of the requirements than any other system previously designed.

Two-Phase A-C Tachometer with Operational Amplifiers

The two basic problems previously discussed are solved in this unit by the use of a 2-phase a-c tachometer with drag-cup rotor, a phase-sensitive detector, and chopper stabilized active operational amplifiers. The drag-cup rotor of the tachometer presents so little inertia to the system that the unit is easily portable and driven from a handheld rubber cone. It may be successfully driven from very small motors without noticeable loading effects. Also, if properly balanced, there is essentially no output voltage with a frequency proportional to speed, and thus no ripple problem other than that of the carrier excitation voltage which may be made appreciably greater than the highest frequency necessary to reproduce the transient cusps. Thus, a very high transient frequency response is possible without interference due to noise or low-frequency ripple proportional to speed.

The use of chopper stabilized active operational amplifiers makes possible the attainment of near perfect differentiation and provides true initial values of acceleration or torque. It is also possible to utilize these operational amplifiers to produce very sharp cut-off filters so that a true response may be obtained nearly up to the carrier frequency. The chopper stabilization of the d-c amplifiers essentially eliminates the drift problem, and the associated recalibration requirements.

System Analysis

Operation

The over-all system operation is best understood by reference to the block diagram of Fig. 1. The 2-phase a-c tachometer with drag-cup rotor is supplied with a 1,500-cps (cycle per second) voltage from a fixed frequency oscillator. The resulting output of the tachometer is a

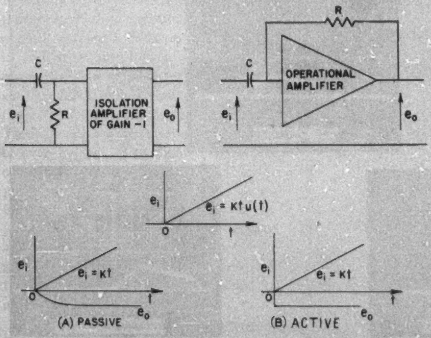

C
R
e_i
ISOLATION AMPLIFIER OF GAIN −1
e_o
C
R
OPERATIONAL AMPLIFIER
e_i
e_o
e_i
$e_i = Ktu(t)$
0
t
e_i
$e_i = Kt$
0
t
e_o
(A) PASSIVE
e_i
$e_i = Kt$
0
t
e_o
(B) ACTIVE

Fig. 4. Active and passive differentiating circuits

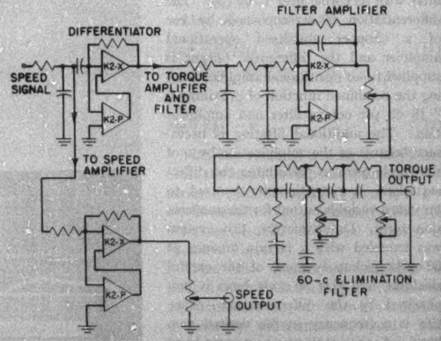

DIFFERENTIATOR
SPEED SIGNAL
K2-X
K2-P
TO TORQUE AMPLIFIER AND FILTER
FILTER AMPLIFIER
K2-X
K2-P
TO SPEED AMPLIFIER
K2-X
K2-P
SPEED OUTPUT
TORQUE OUTPUT
60-c ELIMINATION FILTER

Fig. 5. Simplified operational amplifier circuits

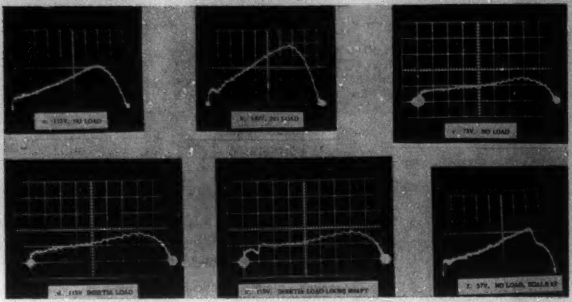


Fig. 6. Torque versus speed, 115-volt, 1-phase capacitor-start motor

1,500-cps voltage with magnitude proportional to speed and a 180-degree phase reversal for reversed rotation. This voltage is fed into a conventional a-c preamplifier. The output of the preamplifier is supplied to a frequency-doubling phase-sensitive detector. Thus the output of the detector is a d-c component of voltage proportional to speed and with polarity indicative of direction of rotation. The ripple voltage on the d-c output has a fundamental of 3,000 cps which is filtered out by an underdamped −12 db (decibel) per octave filter with a sharp corner at 700 cps.

This fairly clean d-c signal is proportional to speed and has a dynamic frequency response somewhat above 700 cps. This signal is split into two channels, one leading to conventional amplification using chopper stabilized operational amplifiers and the resulting adjustable speed signal. The other channel leads to the differentiator. A small amount of additional filtering is accomplished by a conventional R-C (resistance capacitance) filter with its corner near 1,500 cps. The differentiation is accomplished by use of a chopper stabilized operational amplifier and the differential output is supplied to an operational amplifier serving the combined function of providing a −18 db per octave filter and amplification. The additional filtering is necessary because of the amplification by ω of the ripple term in performing the differentiation. A potentiometer is used to provide a variable output for the acceleration term. Unfortunately, this system was troubled with a certain amount of 60-cycle pickup in spite of the careful shielding and this 60-cycle term is emphasized by the differentiation. Since this is a frequency in the desired pass band of the system, a narrow-band parallel-T blocking filter tuned to 60-cycles is used at the acceleration signal output terminals.

All d-c power is supplied from regulated power supplies to provide greater stability for both the oscillator and the operational amplifiers. Provision is made for zeroing the phase-sensitive detector and observing the drift of the operational amplifiers separately from the front panel.

TACHOMETER AND PHASE-SENSITIVE DETECTOR

The tachometer used is a type *10047-2A Eclipse-Pioneer*, 400-cycle 2-phase servomotor with one winding excited from an audio oscillator source of 1,500-cps voltage. The output of the second stator winding is a 1,500-cps voltage with magnitude proportional to speed and 180-degree phase position dependent on direction of rotation. The output is essentially linear with speed to above, 3,600 rpm. A slight quadrature component of voltage is observable at zero speed, but by proper choice of tachometer may be made small enough to have negligible effect.

The tachometer is encased in an aluminum hand-held unit with tapered rubber cone or tip drives; see Fig. 2. A variety of cones or rubber tips may be used to accommodate a range of shaft diameters and center hole sizes.

The 1,500-cps speed-modulated carrier is fed into an a-c preamplifier with adjustable signal level to assure maximum undistorted output for a particular tachometer unit. The output of the a-c amplifier is transformer coupled to a frequency-doubling phase-sensitive detector using silicon diodes. The output of the detector is a d-c signal with magnitude proportional to speed, polarity indicative of rotation, and a fundamental ripple frequency of 3,000 cps. The circuit is shown in Fig. 3. The ripple components are effectively removed by a

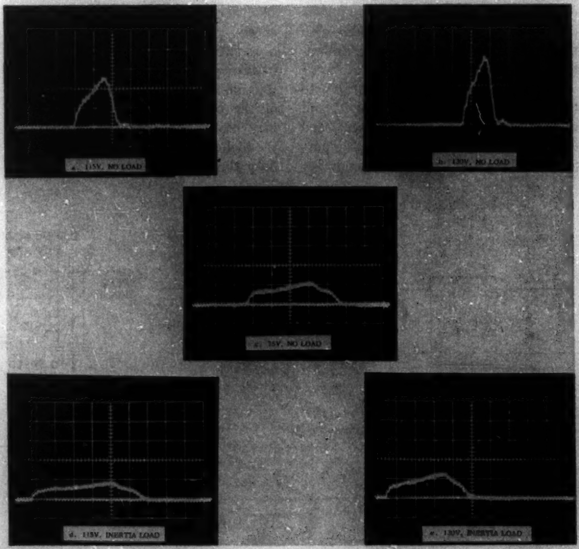


Fig. 7. Acceleration versus time, 115-volt 1-phase capacitor-start motor

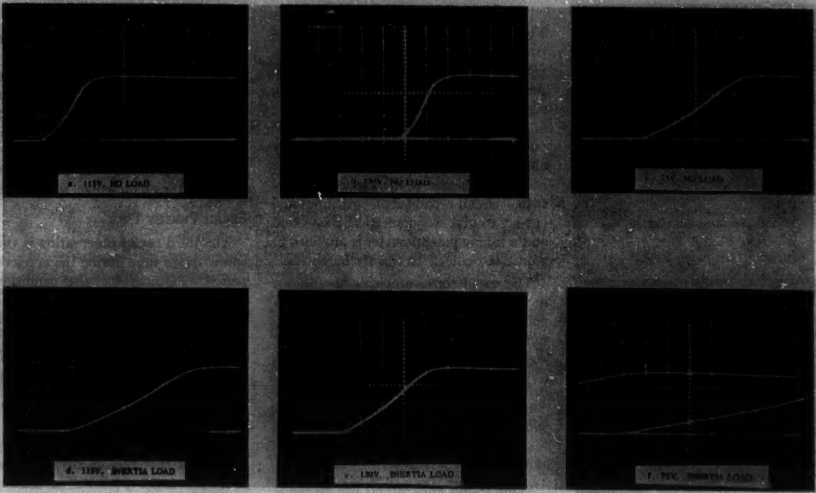


Fig. 8. Speed versus time, 115-volt 1-phase capacitor-start motor

slightly underdamped −12 db/octave R-C filter with a corner frequency of 700 cps. Thus, the speed output signal has a flat pass band up to slightly above 700 cps. A virtual zero speed may be set at any level as well as making adjustment for d-c drift of the detector by a metered d-c null adjustment brought to the front panel. Details of design are straightforward and follow conventional electronic methods.

OPERATIONAL AMPLIFIERS, DIFFERENTIATOR, AND FILTER

The chopper stabilized d-c operationl amplifier is the center of the differentiator, filter, and high-gain d-c sections. It is only by use of these amplifiers that the required initial conditions of the acceleration signal and the absence of d-c drift in both speed and torque outputs can be obtained. The amplifiers used in this unit are type *K2-X* plug-in Philbrick units with type *K2-P* plug-in chopper stabilizers. These have a transient frequency response of over 250 kc. The slight offset of the outputs is easily compensated by scope position adjustment and thus the amplifiers require no zeroing or external adjustment. Three such amplifier sections are used; one for differentiation, the second for amplification and filtering of the acceleration signal, and the third for amplification of the speed signal.

An important problem is the ability of the differentiator to produce an accurate measure of the acceleration while the speed signal is still essentially zero. Consideration will be given to the comparison of a differentiating circuit using an operational amplifier with one using passive R-C components. Consider the two circuits of Fig. 4. Each circuit is subjected to a unit step function of voltage of slope K at time $t=0$.

The passive circuit operational equation is:

$$\frac{e_0}{e_i}=\frac{-RCp}{RCp+1} \quad (1)$$

or

$$e_i=\frac{-1}{RC}\int e_0 dt-e_0=-\left(1+\frac{1}{RCp}\right)e_0 = Kt\mu(t) \quad (2)$$

By differentiating,

$$K=-\left(p+\frac{1}{RC}\right)e_0 \text{ for } t>0 \quad (3)$$

Thus

$$e_0=-KRC+A\epsilon^{-t/RC} \quad (4)$$

If

$$e_0=0 \text{ at } t=0+$$

$$0=-KRC+A \text{ or } A=KRC \quad (5)$$

Thus

$$e_0=-KRC(1-\epsilon^{-t/KC}) \text{ for } t>0 \quad (6)$$

From this

$$\frac{de_0}{dt}=-K\epsilon^{-t/RC} \quad (7)$$

and at

$$t=0+$$

$$\frac{de_0}{dt}(0+)=-K \quad (8)$$

Thus, the passive circuit has a response limited in initial slope to that of the signal. This is plotted in Fig. 4(A).

By comparison, the active circuit equation assuming a perfect amplifier may be written as

$$e_0=-RCpe_i=-RC\frac{d}{dt}(Kt\mu(t)) \quad (9)$$

Thus

$$e_0=-KRC \text{ for } t>0 \quad (10)$$

Also, from equation 9,

$$pe_0=-RCp^2e_i \quad (11)$$

At $t=0+$, it is assumed that $p^2e_i\to\infty$; thus

$$pe_0(0+)=-\infty \quad (12)$$

This has assumed a perfect amplifier. Since the amplifier response is known to be about 250 kc, the actual slope of e_0 at $t=0+$ is limited to approximately 150×10^6 volts per second. A typical response is shown in Fig. 4(B).

Thus, the operational amplifier is limited only by its own frequency response and maximum output values in producing a perfectly differentiated signal. The passive circuit with accompanying amplifier limits the acceleration signal to an

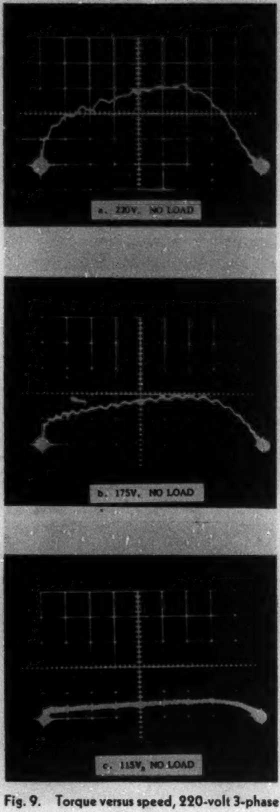


Fig. 9. Torque versus speed, 220-volt 3-phase 3,500-rpm motor

initial slope equal to that of the signal being differentiated for equal voltage output and is incapable of providing an initial torque signal at the moment of zero speed and finite acceleration. A simplified version of the differentiating circuit is shown in Fig. 5.

The torque amplifier and filter shown in Fig. 5 make use of the high input impedance of the operational amplifier together with its feedback properties to produce amplification of the desired signal and an −18 db per octave filtering of harmonics at the same time. The potentiometer adjusts output sensitivity ahead of a parallel-T sharp-tuned filter used to eliminate 60-cycle pickup. The twin T circuit is padded on both input and output to give a symmetrical pass band and a tuning potentiometer is provided to adjust the cut off to exactly 60 cycles.

The speed signal amplifier also shown in Fig. 5 has a fixed gain of 10 with a potentiometer output for variable sensitivity. It is a straightforward chopper stabilized operational amplifier in all respects.

Results

SPEED VERSUS TORQUE

Many useful applications can be made of the instrument described in the preceding paragraphs. The three more apparent ones are the observation of speed versus torque or acceleration, acceleration versus time, and speed versus time. The recording of speed versus acceleration will be considered first. Typical results of the unit are shown in Fig. 6 by use of pictures of the instantaneous trace displayed on an oscilloscope. These pictures were recorded while the motor was undergoing a normal starting acceleration with various voltages and external loads. As such, they represent the transient torque versus speed relationship at the shaft of the motor. Values of initial torque, maximum torque, full-load speed torque are easily observed and may be correlated with the linear speed co-ordinate. A switching cusp is apparent in Fig. 6. It is also possible to observe the acceleration versus speed relationship of a loaded motor under normal starting conditions. The curve of Fig. 6 is for an inertia load and demonstrates the net torque available for acceleration.

ACCELERATION VERSUS TIME

A second useful result is the observation of the acceleration versus time of a motor either unloaded or loaded during its starting period. Fig. 7 demonstrates the acceleration versus time relations for most of the acceleration versus speed plots of Fig. 6. A linear calibrated time axis is produced by the internal sweep circuit of the oscilloscope; see also Fig. 9.

SPEED VERSUS TIME

The third useful observation is that of speed versus time during the acceleration period. Pictures for the condition of Fig. 6 are shown in Fig. 8. These results give the time necessary for the motor to attain any fraction of full speed under any desired load condition. Again a calibrated linear time scale is supplied by the oscilloscope.

Conclusions

It is felt that this instrument provides the first truly portable instantaneous speed-torque recorder to the motor manufacturing industry. Its stability and simplicity of operation, combined with its ease of application, make it equally useful in the laboratory or on the production line. With this instrument it is possible to determine in a matter of seconds the complete torque-speed characteristic including cusps due to harmonics or switching and to obtain this without the motor being loaded. Thus, it is a vital tool in the experimental verification of new or modified designs, in production line quality control, and for on the spot trouble shooting. Only with the recent availability of stabilized operational amplifiers and high precision servo components has such an instrument been feasible at a reasonable price.

References

1. SPEED-TORQUE CURVES ON AN OSCILLOSCOPE, S. Chang. *Electrical Manufacturing*, New York, N. Y., Sept. 1950, pp. 452–56.
2. INDUCTION MOTOR TORQUE AND ITS MEASUREMENT, E. W. Henderson, K. M. Chirgwin. *The Engineering Journal*, Montreal, Que., Canada Nov. 1953.

The Relative Merits of Incremental and Conventional Digital Computers in Air-Borne Real-Time Control

S. M. SHACKELL
NONMEMBER AIEE

J. G. TRYON
NONMEMBER AIEE

IN THE conventional or general-purpose digital computer, operands are tranferred in their entirety from one functional block to another. The inputs and outputs of such a block are simply the values of the constants and variables concerned. In the incremental computer, however, only portions of the operands are transferred. The inputs and outputs of a functional block are increments of the quantities concerned; the digital differential analyzer is one example.[1]

In these definitions, a functional block is an operation, e.g., multiplication or limiting. An operand is one of the quantities being calculated or entering into a calculation, i.e., a multiplier, or a quantity to be limited.

The purpose of this paper is to evaluate the relative merits of the two designs in air-borne real-time control. These designs are also used extensively in industrial process control. However, the experience of the authors is in air-borne systems, and the opinions expressed here will be confined to this field.

In real-time control, the computer guides some physical system along a prescribed path. The inputs to the computer are the positions, velocities, and other quantities that describe the state of the system; the outputs are signals that control the system. An autopilot is an example. Here the system is the aircraft, and the computer outputs control ailerons, elevators, and rudder.

First, observe that the two designs are related, differing in the way in which changes in variables are handled, and that both have sufficient potential capability. After defining the air-borne real-time control problem, this study will compare actual computers, concluding that each machine has its place, and that the place of the incremental is limited.

Kinship Between Incremental and Conventional Computers

In the first incremental machines increments could have only two values +1 and −1. In later "ternary" machines, improvement was obtained by increasing the number of possible values to three, +1, 0, and −1. A dual-mode incremental computer has been considered in which the number of different values is increased still further to five, +32, +1, 0, −1, −32. It is obvious that the number may be increased still further until the increments may take on all possible values that operands may attain.

If more and more possible values of increments are allowed, changes in operands can be transferred in fewer steps. It is obvious that if increments can have all possible values of any number in the machine, any change can be transferred in a single step. Such a computer is merely a conventional computer.

From this it may be seen that the general-purpose computer and the digital differential analyzer are the opposite extremes of a single class of computers. Indeed, the conventional design is the limiting case of the incremental design.

An Essential Difference

Much understanding may be gained by studying an essential difference between the incremental and the conventional machines. A common characteristic of all real-time control computers is that they have inputs and outputs which are functions of time. As the inputs change with time, the computers cyclically recalculate correspondingly changed values of the outputs. Furthermore, as their calculations proceed, all computers evaluate intermediate quantities. Although these intermediate quantities are necessary to the calculation of the outputs, they are not outputs.

The essential difference is that, in the incremental computer, the intermediate quantities and outputs are retained from cycle to cycle and brought up to date in each cycle. Changes in the inputs since the most recent calculation are propagated through the calculation as changes in the intermediate quantities to become changes in the outputs. In the conventional computer, on the other hand, the intermediate quantities and outputs of one cycle are discarded and have no use in the next recalculation. Each quantity is recomputed afresh in each cycle; there is no essential holdover from previous cycles. (Certain exceptions will be discussed.)

The two computers are very like two people of different temperaments who are given new values to plug into a set of formulas with which they have just been working. One (the conventional) tears up the old work sheet in disgust and starts over. The other (the incremental) patiently computes the change required in the first entery on the work sheet, erases the old entry, puts down the new, and continues on in the same fashion across the page.

For example, suppose that the calculation is $q=wxy+z$, where w, x, y, and z are inputs, and q is the output. Let (wx) be the intermediate quantity and let i indicate the cycle. The conventional computer might use the computation equations

$$(wx)_i=w_ix_i \tag{1}$$

and

$$q_i=(wx)_iy_i+z_1 \tag{2}$$

The incremental might use

$$\Delta w_i=w_i-w_{i-1}$$

$$\Delta x_i=x_i-x_{i-1}$$

$$\Delta(wx)_i=w_{i-1}\Delta x_i+x_i\Delta w_i$$

$$(wx)_i=(wx)_{i-1}+\Delta(wx)_i \tag{3}$$

and

$$\Delta y_i=y_i-y_{i-1}$$

$$\Delta(wxy)_i=(wx)_{i-1}\Delta y_i+y_i\Delta(wx)_i$$

$$\Delta z_i=z_i-z_{i-1}$$

$$\Delta q_i=\Delta(wxy)_i+\Delta z_i$$

$$q_i=q_{i-1}+\Delta q_i \tag{4}$$

Compare equations 1 and 2 with equations 3 and 4 respectively. Note that values from the $(i-1)$th cycle do not

---

Paper 60-120, recommended by the AIEE Computing Devices Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., January 31–February 5, 1960. Manuscript submitted November 2, 1959; made available for printing January 29, 1960.

S. M. SHACKELL, deceased, was with Stevens Institute of Technology, Hoboken, N. J., and J. G. TRYON, formerly with the Bell Telephone Laboratories, Inc., Whippany, N. J., is with the University of Alaska, College, Alaska. The work was supported by the U. S. Air Force, Air Material Command, and was a part of the TRADIC Computer Research Project at Bell Telephone Laboratories. The opinions presented are those of the authors and do not necessarily reflect Air Force opinions or attitudes on the subject.

appear in the conventional equations. Note that in the incremental equations, the input changes Δw_i, Δx_i, Δy_i, and Δs_i are propagated through the calculation. They first appear as a change $\Delta(wx)_i$ in the intermediate quantity wx, and later as a change Δq_i in the output q.

Calculations where conventional computers do not start afresh in each cycle are the exceptions already mentioned. They are only partial exceptions and do not destroy the difference just observed since in each the phenomenon of propagation of changes in inputs through changes in intermediate values to become changes in outputs is missing.

Some of the exceptional calculations are integration and differentiation with respect to time, servoed solutions (a form of successive approximation), and prediction.

For example, in time integration, both computers approximate the integral

$$s=\int_0^t y dt \tag{5}$$

by

$$s_i=s_{i-1}+\Delta s_i \tag{6}$$

and

$$\Delta s_i=y_i T \tag{7}$$

where i indexes the cycle and T is the period. Here the conventional design by the very nature of things must go back to the previous value of the integral s_{i-1} to compute the present value s_i, and so is to some extent incremental. However, it does not go back to the inputs or intermediate quantities of the previous cycle in computing y_i, and so the calculation is emphatically not an incremental one.

In summary, it may be said that in real-time control: 1. The incremental computer retains previous intermediate quantities and results and corrects them to allow for changes in the inputs (it does this by propagating the changes through the calculation), and 2. the conventional computer discards previous intermediate quantities and outputs, except where they are necessary to the calculation. There is no propagation of changes.

Logical Capabilities of Incremental Computers

Incremental computers often are considered to have very limited logical capabilities but this is not necessarily true. In the matter of abstract logical design, what the conventional computer can do, the incremental can also do. The capabilities of economical machines are quite different.

Historically speaking, the incremental computer evolved from a rather restricted design. In the original digital differential analyzer, every calculation was reduced to integrations (and additions), which in turn were represented by finite summations. As difficulties were found, the design evolved until now the incremental is far more than a digital version of a differential analyzer. Integration is no longer the primary operation. Rather, the type does arithmetic in a particular, incremental fashion. This arithmetic is just as valid as that done in a conventional digital computer; only the method is unusual.

EXACT INCREMENTAL COMPUTATION

The belief that incremental computers are not exact is based on the awareness of certain weaknesses of the early models. Actually, incremental calculation can be made as exact as may be desired.

First, incrementals may be equipped with exact algorithms. Indeed, multiplication has long been exact; see equations 3. Division is usually done by servoed multiplication, and is therefore not exact. It may, however, be done as follows. To solve $q=x/y$, use the incremental equations:

$$y_i=y_{i-1}+\Delta y_i$$

$$r_i^*=r_{i-1}+\Delta x_i-q_{i-1}\Delta y_i$$

$$\Delta q_i=\left[\frac{r_i^*}{y_i}\right]_{\text{rounded}} \tag{8}$$

$$r_i=r_i^*-y_i\Delta q_i$$

The solution must be started with

$$q_0=\frac{x_0}{y_0} \text{ and } r_0=0 \tag{9}$$

It is then exactly equivalent to the division equation of the conventional machine:

$$q_i+\frac{r_i}{y_i}=\frac{x_i}{y_i} \tag{10}$$

where r_i is the remainder. Here r_i^* is a temporary remainder, i indexes the cycle, and subscript zeros indicate initial conditions.

To achieve exact calculation, it is also necessary that the increments be able to take on a variety of values. If each successive output value is to be exact, the computer must be equipped to handle any change in value that may occur. The full change must be processed to full precision.

Incremental computers have never been constructed with this feature. The necessary extension is perfectly straightforward in theory. Instead of a machine in which increment values are restricted to +1, 0, −1, a design must be provided in which: 1. the largest permissible increment is larger than any expected change in any input, intermediate value, or result, and 2. increments are expressed with a sufficient number of digits to achieve the desired accuracy.

The dual-mode incremental computer already mentioned marked the beginnings of such a design. In it increments could take on the values +32, +1, 0, −1, −32. A better set of values would be +32, +31, +30, +29, . . ., +2, +1 0, −1, −2, . . ., −31, −32. Such increments would be expressed as 6-digit binary numbers.

It should be noted that an exact incremental must do actual multiplication and division in the course of the algorithms. Multiplication occurs in the formation of $w_{i-1}\ \Delta x_i$, $x_i\ \Delta w_i$, $q_{i-1}\ \Delta y_i$, and $y_i\ \Delta q_i$. Division occurs in

$$\Delta q_i=\left[\frac{r_i^*}{y_i}\right]_{\text{rounded}}$$

These multiplications are not simple gate-and-add operations as in the digital differential analyzer. They are multiplications of, for example, 16-bit numbers by 6-bit numbers. Similarly, division requires that the quotient be carried to perhaps six bits.

Exact computation also requires that accurate equations be used. It has been customary in digital differential analyzers to evaluate many functions by integrations, where of course the integrations are approximate by finite summations. The classic example is the computation of sine and cosine which is done by the cross-coupled integrations

$$\sin\theta=\int\cos\theta d\theta \tag{11}$$

and

$$\cos\theta=-\int\sin\theta d\theta \tag{12}$$

Much work has been put into this method and it can be very satisfactory. However, it can also be quite inaccurate, i.e., if the increments are too large.

Exact evaluation may be obtained by expanding the function in an appropriate series, e.g., a power series. This series will be made up of multiplications, divisions, additions, and subtractions and so can be evaluated exactly using the algorithms already suggested.

It should be obvious from the above that an exact incremental computer is an expensive extension of the old digital differential analyzer. Cost considerations, however, should not obscure the present issue. There is nothing inherent in the logical principle of incremental calculation that limits accuracy; the incremental can be just as exact as the conventional computer.

Because of finite word length, increment size, and computation rate, there is, of course, no such thing as a perfectly exact computer, conventional or incremental. In the interest of simplicity, this point has not been emphasized in the foregoing. A review of the analysis will show that in every respect, however, the accuracy of an incremental design may be made identical with that of a conventional design.

DECISIONS

Real-time control involves decisions e.g., when an output must be limited or when one calculation must be substituted for another. Such decisions have been rare in incremental designs. There are two ways, however, in which they can be handled, one corresponding to address modification in the conventional machine, and the other to transfer of control.

A decision element or limiter has appeared in several incremental designs. Straightforward generalization suggests a functional block, called selection, corresponding to address modification. The block is essentially a single-pole double-throw switch. It is designed to select between the changes in two quantities just as address modification in a conventional machine may select one of two quantities. The selection is made on the basis of the sign of a third quantity, the criterion. Stated exactly, the block performs

$$q_i = q_{i-1} + \Delta q_i$$

$$\Delta z_i = \Delta x_i \text{ if } q_i \geq 0$$

$$= \Delta y_i \text{ if } q_i < 0 \qquad (13)$$

Here, Δz_1 is the output increment, and Δx_1 and Δy_i are the input increments between which selection is made. The quantity q_i is the criterion on which selection is made; it is kept up to date by addition of the third input which is Δq_i. As before, i indexes the cycle.

Selection would typically be used in changing of the value of a coefficient in the calculation upon the change in sign of a key quantity. Two servo functional blocks and a selection block would be required.

Regarding equipment cost, the selection operation is no more expensive than other operations. However, its use may involve expensive delays. Whenever the selection action triggers off a large change in some quantity, a slewing time must elapse before the change is complete. The slewing time is determined by the magnitude of the change and the largest increment possible in the machine; it may be long or short.

The foregoing shows how the incremental computer may accomplish the equivalent of conditional address modification, and thus substitute one number for another. Where one calculation must be substituted for another, the incremental needs a conditional jump operation analogous to that of the conventional machine, e.g., in an autopilot it might be necessary that a different program be used above a specified altitude.

To accomplish a conditional jump, the incremental must: 1. sense that a jump is needed, 2. reload all registers with initial values associated with the new program, and 3. switch to the new program of instructions. Reloading may be done by the regular loading circuits, which must read from the initial-value bands on the drum that are appropriate to the new program. Switching to the new program is done by switching reading heads on the drum. The process is sequenced by a flip-flop circuit which is set in action by the condition calling for the jump. This may be a change in the sign of some criterion, or it may be an intervention by the operator.

Equipment cost for the scheme includes extra initial-value and program tracks. There must be the sequencing circuit, and sufficient equipment to detect the need for the jump. Equipment cost is probably not too high; the serious cost is in slewing time. After every jump there will be a hiatus during which quantities work their way from initial to up-to-date values.

One way to reduce slewing time is to provide large increments, as has already been suggested. Another is to maintain all calculations simultaneously, switching only the final outputs. A third is to provide a conventional computer that supports the incremental by providing it with up-to-date initial values. All are expensive.

It may be noted that the "red-tape" decisions of conventional computers have no counterpart in incrementals. This is because subroutines and loops are impractical in incremental calculation.

If both forms of decision are provided, the authors believe that the incremental computer can, at least in principle, make all decisions that conventional computers can make.

CORRECTION OF ERRORS

At first glance it appears that correction of errors is difficult or impossible in an incremental computer. Since every calculation is of the form

$$x_i = x_{i-1} + \Delta x_i \qquad (14)$$

it appears that an error in the present calculation propagates into all future calculation and can never be wiped out. Actually, the appearance is not necessarily true. Once the computer has been provided with the conditional jump facility just described, it can correct any error that it can discover. This places it on a par with the conventional machine.

The proposed scheme is as follows: Error-detecting checks will be maintained. On receiving an alarm from one, the conditional jump facility will be used to reload initial conditions. After sufficient time has elapsed to cover slewing, the calculation should be in good shape again.

All the standard error-detecting checks are possible. Parity checks on the storage, reading, and/or transmission of both full quantities and increments may be provided; also overflow checks on the magnitude of quantities may be provided. Tests for the continuity (smoothness) of quantities are easy, since the rates of change of all quantities are available as the increments. Arithmetic may be checked by providing duplicate circuitry or by repeating the calculations.

The cost of this scheme is in the check circuitry and the jump facility. The slew time which must elapse before an error is repaired may be significant. The cost will be large or small depending on the thoroughness with which errors are detected.

The operation of the scheme requires two arrangements which have been a part of incremental design for some time. The first is the permanent recording of the initial conditions; the second, the provision for what is called "no loss of reference," i.e., bringing into the computer a complete new value of each input in each cycle. These new values are compared with the computer's versions, and the changes in the inputs are calculated accordingly.

SUMMARY

It appears that as far as logical capabilities are concerned, the incremental design is a very powerful one. Indeed, it seems that two computers, one incremental and the other conventional, may be externally equivalent. Both can produce the same outputs when fed the same inputs.

The choice between the arithmetic and incremental principles cannot therefore be made on the basis of capability of the logical design; it must rest on equipment cost, maintainability, convenience in programming, and similar factors.

Real-Time Control Problem

In air-borne real-time control, some systems must be steered or otherwise directed in response to physical stimuli. Action must be fast enough and accurate enough to achieve the required degree of control, whatever it may be.

The work of the computer begins with the conversion of the inputs. Typically, these are continuous physical quantities such as air speed, acceleration, and target range. Each must be sampled repetitively and thus converted to a sequence of digital values.

The problem continues with the actual calculation, which is a cyclical affair. The same calculation is repeated again and again, once for each new sampling of input values. The results are the outputs. These appear first as sequences of digital values and then, after conversion, as continuous physical quantities.

The problem concludes with this conversion. The output sequences are converted to continuous physical signals, like shaft positions, which in turn control the physical system.

For present purposes, attention should be concentrated on the calculation itself. Input and output conversions are difficult; however, they tend to have the same realizations in both designs. Ever since the addition of the no-loss-of-reference feature to the digital differential analyzer, there has been little difference between the input-output equipment of the two types.

The typical calculation is made up of co-ordinate conversion, smoothing and prediction, time integration, evaluation of arbitrary functions (usually approximated by polynomials), and general calculation. These computations, which involve algebra, trigonometry, and calculus, are realized by programs of arithmetical operations. In the case of the conventional computer, the operations are those of ordinary arithmetic; in the case of the incremental, they are those of the special incremental arithmetic.

Besides arithmetic, the calculation sometimes involves decisions (conditional transfers of control). For example, a bombing calculation must determine the moment for dropping the bomb. Such decisions are not frequently encountered within a single individual calculation; they are, however, common in the routines which assemble individual calculations into a system and in routines for control of error.

It is noteworthy that the calculations deal with quantities which are continuous functions of time. Not only are the functions continuous, but their derivatives are bounded by physical limitations.

Capabilities of Actual Computers

The capabilities of typical machines will now be compared. To be specific, consider: 1. a simple incremental, 2. an incremental with provision for fast slewing, and 3. a conventional computer. Here "typical" means well proportioned for the class of control jobs for which the type is suited. No attempt will be made to compare equivalent computers, for to do so would be unfair to at least one. The types are suitable for quite different jobs; to consider equivalent computers would be like trying to decide the relative merits of a baseball team and a football team on the basis of their showings at soccer.

This discussion continues to be limited to air-borne computers. Numerical estimates are given where possible, in order to indicate orders of magnitude. They should be correct to within a factor of 3. They are based on transistor circuitry and on magnetic-drum storage in the case of the incremental computers and core storage in the case of the conventional (all circa 1956).

The simple incremental is the incremental with increment sizes limited to +1, 0, −1. It is usually called a digital differential analyzer. The incremental with provision for fast slewing is a machine with multiple increment sizes, e.g., +32, +1, 0, −1, −32.

SPEED

The calculation must be done with sufficient speed. Speed has two aspects: First, it is necessary that the initial solution of a new problem be delivered with reasonable promptness. This requires that large changes in variables be processed in each computation cycle, since the computer must move promptly from whatever state it finds itself in to the state demanded by the problem. (The time required for such motion is known as "slewing time.") Second, the problem must be solved frequently enough to keep up with whatever high-frequency changes in inputs occur. This calls for frequent recalculation rather than for the handling of large changes. (Frequency of recalculation is measured by calculation or iteration frequency. It determines the bandwidth that the computer can handle which is in practice about one eighth of the calculation frequency.)

The incremental computers typically have the better computation frequency. Both simple and fast-slewing incrementals deliver as many as 200 solutions/sec (per second). Contrast this with the 10 or 20 solutions/sec that are delivered by the typical conventional machine. The incrementals are an order of magnitude better. They are satisfactory in a control system requiring a bandwidth of about 25 cps (cycles/sec), whereas the conventional computer can work up to only about 2.5 cps. The incremental machines gain this advantage by the efficiency of their arithmetical operations, which involve many fewer elementary logical operations. This efficiency is so great that the circuitry of an incremental moves at a very lazy rate. Both types could be speeded up by using more or faster equipment, but the relative advantage would remain with the incremental.

With regard to slewing time, however, the conventional machine is the better computer. Starting from any state, it can produce the first correct result in one computation time, or about 0.1 or 0.05 sec. The simple incremental may require 300 sec for the same achievement. The incremental with fast slewing needs perhaps 10 sec. The conventional machine is faster by two or three orders of magnitude, an impressive advantage which is rooted in the basic logical principle of the conventional computer. It recomputes each quantity anew in every cycle, and so can establish completely new values of all quantities in one cycle of calculation.

The poor showing of the incremental computer is also rooted in the basic logical design. The incremental propagates changes in its variables through the calculation, and gains its impressive simplicity by sharply limiting the magnitude of the changes which can be made in each cycle. However, by adding more equipment, as in the incremental with fast slewing, the increments can be made larger and the disadvantage reduced. The cost is high, and our work suggests that the cost of a conventional machine is reached long before the slewing time of the machine is approached. (Workers on the West Coast have used two computers in the same system; a conventional machine to handle slewing and an incremental to handle the continuing calculation. This has promising possibilities.)

In another view of the matter, it may be observed that the incremental machine is rate-limited. Although it can process a great many changes in a quantity per second, the changes are quite small. Conventional machines, on the

Vol.
Pt.
196

other hand, are frequency-limited. Although they can process very large changes in a quantity, they do so only with limited frequency.

Considering still another aspect of speed, it may be said that the delay in a conventional computer is a constant. Being of the order of the computation period, it is perhaps 0.05 sec in a fast computer. The delay in an incremental machine is a function of the magnitude of the changes being processed. It is as low as 0.005 sec (the iteration period) when changes are small; it may be up to 10 or even 300 sec (the slewing time) when changes are very large.

LOGICAL AND MATHEMATICAL CAPABILITIES

Although incremental and conventional designs have the same abilities in principle, their logical and mathematical abilities are very different in practice. The difference arises from the following facts:

1. Less time and equipment are required by incremental machines to carry out repetitive algebraic and trigonometric calculations on continuous quantities. Time integrations and the evaluation of polynomials are natural operations for the incremental.
2. Far less time is required by the conventional machine to implement decisions.

With regard to the first point, in repetitive calculation, the incremental computers capitalize on the fact that real-time control deals with continuous quantities. They work with the changes in these quantities, rather than the entire quantities, thus saving both time and equipment, for the changes are of limited magnitude and the operations are thereby simplified. The conventional computer, however, is fundamentally unable to capitalize on continuity, since it recomputes all values in each cycle. It takes more equipment to do the same calculation, as well as more time.

On the other hand, conventional computers implement decisions promptly. Since each new cycle of calculation produces new values of all quantities, only one computation time is needed to get started on a new path. As already mentioned, this time may be as low as 0.1 or 0.05 sec. The incrementals are slow in implementing decisions because they must drop the old calculation and start up the new one. To do so requires going from an arbitrary state (the state of performing the old calculation) to a different arbitrary state (the new one). In the course of the change, each quantity must work its way to the new value, one increment at a time. To do so may require one slewing time, which may be 10 to 300 sec depending on the amount of equipment that has been spent on fast slewing.

The incrementals have no difficulty in making a decision; this is done promptly and cheaply by a selection block. Their difficulty lies in getting started on the work required by the decision. If it is desired to improve the incrementals in this respect, one may maintain all possible calculations simultaneously in the computer and implement the decision by switching outputs, which is an expensive process. The conventional design, therefore, has the advantage when decisions requiring appreciable changes in the calculation are to be made; the incremental with fast slewing comes next; and the simple incremental is a poor third.

ACCURACY

Both designs work to the same accuracy in real-time control. There is no difficulty in obtaining the needed accuracy, and no gain in providing more.

As accuracy is increased, the computation time of the conventional computer increases more sharply than that of the incremental. This is because multiplication time in the conventional machine tends to go as the square of the word length, while in the incremental it goes about as the first power.

GENERAL INFORMATION PROCESSING

Real-time control does not now include general information processing (by which is meant nonalgebraic nontrigonometric work such as sorting, collating, logical analysis, and the processing of "noisy" data of low correlation). However, large real-time control systems are beginning to require a tremendous amount of this sort of work. Accordingly, it is noteworthy that the conventional design can handle general information processing. It is not very efficient, but it can do the job.

The incremental computer is quite unable to handle this type of calculation. Some basic change in principle would be necessary before it could complete such a job in acceptable time. "Noisy," uncorrelated data cannot be accepted by the internal registers of an incremental; sorting and logical analysis are equally impossible. The difficulty is that all the data-processing operations involve large changes in the content of registers in every cycle of calculation, and these the incremental cannot achieve.

Physical Characteristics

In describing physical characteristics the discussion will continue in terms of a simple incremental computer, an incremental with fast slewing, and a conventional computer. Discussion is still limited to air-borne problems.

By physical characteristics is meant weight, size, and number of parts. Weight and size determine some of the indirect costs of the computer. All three factors tend to determine the manufacturing cost.

The simple incremental machine is lighter and smaller than the conventional machine. It is perhaps only half as heavy and half as voluminous. The two reasons for this difference are that: 1. the incremental is efficient in the simple problems for which it is suited, and 2. the conventional, being the one which can handle the larger, more involved problems, is assigned the bigger problems. Thus, since it does more, it is larger and heavier.

The sinple incremental is also less complex, having about one quarter as many individual parts. The same two reasons again apply, and a third, very powerful, factor appears. The incremental is built around a magnetic drum which it uses with great efficiency. Furthermore, the logical circuitry is highly efficient, much of it being used for several purposes.

The characteristics just described apply to the plain incremental computer with the very high slewing time, in which increments have the values +1, 0, −1. If the more complicated incremental computer with better slewing time is considered, the weight, volume, and number of parts approach those of the conventional design.

The most striking physical characteristic of incremental machines is that they all use a magnetic drum. They are built around the drum, gain their advantages from it, and would be quite impractical without it. It is conceivable that an incremental machine could be built around some other storage device; indeed, one of the authors has started a machine using magneto-strictive delay lines.

Designability and Maintainability

The description now turns to those characteristics which, although nonphysical, are by no means imaginary. Designability and maintainability are very real items in the description of a computer; the former determines much of the cost of building the machine; the latter fixes the cost of running it.

In this area, opinions vary, but the matter is of sufficient importance that consideration is necessary, and for this reason the authors' opinions are included here.

Incremental designs are not clean-cut, and are difficult to understand and explain. Many of their features and much of their economy are obtained by low cunning. It is difficult to predict their behavior in dynamic situations. Their designability is therefore not considered good.

The conventional design is also not clean-cut but it is more so than that of the incremental. While it is difficult to understand and also is tainted with low cunning, it is, the authors believe, not so much so as the incremental. Its behavior in dynamic situations can be predicted somewhat more easily since one need not allow for rate-limiting; therefore, the designability is somewhat better.

More significant than original designability is the designability of changes. Often, and indeed, usually, when a real-time computer is more or less finished, it is necessary to change the design to allow for some change in the system in which the design is employed. Such changes may even have to be made after production is started. Here the conventional machine is much more flexible and much easier to work with.

When maintainability is considered, the authors believe that even the simple incremental must again take second place to the conventional. Although it has fewer parts, it is harder to troubleshoot. There seems to be no basic reason for this; however, as commonly designed, incremental machines are untidy and inconvenient to work on. In particular, the extensive time-sharing of circuits that makes them so compact also makes them hard to troubleshoot.

There seems to be no reason why incremental computers cannot be designed for easy maintenance; however, it is to be expected that a good deal more equipment would have to be added. It is important to differentiate between maintainability and reliability: maintainability measures the ease with which an existing fault may be found; reliability measures the rate at which faults are generated.

If the incremental computer has poor maintainability, so does the conventional. However, the conventional does seem to be easier to repair than the incremental because: 1. step-by-step operation is easy, 2. a break-point facility is easy, 3. time sharing does not complicate the circuits excessively, and 4. diagnostic programs may be conveniently used.

Reliability

Reliability means the ability of an equipment to perform satisfactorily throughout a specified period of operation, provided that it is in good shape at the start. Two kinds of trouble occur in any digital computer. One is malfunction, the transient trouble that appears, does its damage, and departs, leaving behind it an error which may or may not propagate. The other type of trouble is the failure, the permanent trouble, e.g., a burned-out transistor.

The simple incremental computer has fewer failures because of its fewer components. (It should be remembered that it needs fewer components, in part because it tackles easier problems, and in part because it uses its components very efficiently.) It is reasonable to guess that the typical simple incremental machine has only one quarter as many failures as the typical conventional machine. This is a large advantage. The incremental with fast slewing has almost as many parts as the conventional, and so will have as many failures.

It is observed that the simple incremental computer also has the fewer malfunctions. However, the errors from these malfunctions are "devilish." Because of the basic nature of an incremental each error remains indefinitely. This, it will be remembered, is because the incremental carries over old values into new calculations. Even so, perhaps one half of the malfunctions cause negligible trouble because they affect only one increment, because they occur in an unimportant calculation, etc. The remaining half, however, spoil the performance.

In the conventional computer, by contrast, most malfunctions cause no more than a temporary inconvenience. A check of one computer, for instance, indicated that only one tenth of all its malfunctions would cause serious errors. This is because the machine recomputes almost all values afresh in each cycle, and it is only where the problem being solved involves carry-over of information from one cycle to the next that errors are not wiped out promptly.

The net result is, apparently, that the simple incremental and conventional computers are about alike in the number of performances spoiled by malfunctions. The incremental computer with fast slewing is in a much poorer position. It has almost as much equipment as the conventional and therefore has almost as many malfunctions, which all have lasting effects, and about half of them will be serious enough to spoil performances.

Thus far computers without special error-correcting facilities have been discussed. The best error-control schemes known to the authors are based on programmed detection and correction, rather than on the use of redundant equipment, which is prohibitively expensive. These schemes detect errors by mathematical means, and correct them by repetition of the calculation. It is reasonable to expect that they will catch and deal with two thirds of all malfunctions.[2]

Now, it is easy to add programmed error control in the conventional computer. A few extra operations, perhaps 100, must be added to the calculation to be checked. It is equally easy to add the equivalent control program to either incremental. The additional operations are added in straightforward fashion. However, before the control program is effective in an incremental, short slewing time must be provided. The control program, to be effective, must be able to discard the faulty solution and slew in a new one from the permanently recorded initial values, all in a short time. Hence, it is impractical to use an error-correcting program in the simple incremental.

It thus appears that, with respect to calculations lost because of malfunctions, the best machine is the conventional computer with error control. The conventional computer without error control, the simple incremental without error control, and the incremental with fast slew and error control follow close behind. The incremental with fast slewing but no error control places far behind the others.

To arrive at a picture of over-all reliability, both failures and malfunctions must be considered. Malfunctions occur about 10 or 20 times as often as failures, so the malfunctions dominate and the ranking just given holds for over-all reliability.

In the following it will be assumed that the conventional computer and the incremental with fast slewing are provided with error-control programs. To build either machine without error control would be poor design. That the simple incremental has no such control will be assumed.

Comparative Evaluation

It is clear that no one of the designs is superior to the others. Rather, each is suitable for a different class of problems. The air-borne problems for which each design is appropriate will now be described.

PROBLEMS FOR THE SIMPLE INCREMENTAL COMPUTER

The incremental computer with limited increment sizes (and hence long slewing time) has these advantages:

1. Very high computation frequency.
2. Very good capacity for repetitive calculations upon continuous quantities.
3. Small volume, low weight, fewer elements.
4. Excellent reliability with respect to failures.

Balancing the advantages are these disadvantages:

1. Very long slewing time.
2. Poor ability to implement decisions or do general information processing.
3. Not very good designability and maintainability.
4. Adequate reliability with respect to malfunctions.

This design is therefore suitable for the elementary real-time control problems, e.g., in an autopilot. Such problems can use all of the advantages of the simple incremental and, indeed, may demand some of them. The computer which rides in and controls a missile, for example, will unquestionably require small volume, low weight, and excellent reliability with respect to failures. It may also require very high computation frequency for stability.

The elementary problems involve no general information processing. Decisions are few and involve no abrupt changes in the calculation, so the weaknesses of the incremental in this regard are immaterial. Any deficiencies in designability and maintainability can be overlooked. Reliability with respect to malfunctions is satisfactory.

However, the simple incremental cannot be used where fast slewing is required. This eliminates any problem in which the computer gets little warning, and in which it must produce results immediately after the first data are supplied. The simple incremental is therefore a poor choice for defensive fire control in a bomber, for example, where only a few seconds may be allowed for engaging the target. This is regrettable, since its high computation frequency is exactly what is needed for fast response after the target is engaged. It is, however, at no disadvantage in handling the missile because it can be started on the problem ahead of time.

The simple incremental is not suitable for compound problems in which first one and then another calculation must be undertaken.

PROBLEMS FOR INCREMENTAL COMPUTER WITH FAST SLEWING

When the incremental is provided with multiple increment sizes so that slewing time is reduced at the price of more equipment, the result has these advantages:

1. Very high computation frequency.
2. Very good capacity for repetitive calculations upon continuous quantities.

The result has an intermediate position, neither very good nor very poor, with respect to many of the other factors, specifically:

1. Moderate volume, weight, and number of elements.
2. Satisfactory reliability.
3. Acceptable slewing time.
4. Sufficient ability in implementing decisions to handle a few.

Clearly, its disadvantages are:

1. Poor ability in general information-processing.
2. Poor designability and maintainability.

The design, therefore, is appropriate where very high computation frequency is needed for dynamic response and where general information processing is not necessary. A suitable problem appears to be defensive fire control, which requires both high computation frequency and short slewing time.

However, since this design is weak in general information-processing and can handle only a few decisions in reasonable time, it is not suited for compound problems. The weaknesses in its designability and maintainability also limit its usefulness.

PROBLEMS FOR CONVENTIONAL COMPUTER

The conventional design is strong in these points:

1. Very short slewing time.
2. Excellent ability to implement decisions and to do general information-processing.
3. Excellent designability.
4. Better reliability.

It is entirely satisfactory in these qualities:

1. Adequate capacity for repetitive calculations upon continuous quantities.
2. Reasonable volume and weight.
3. Reasonable maintainability.

Its weaknesses are:

1. Moderate computation frequency.
2. Larger number of elements.

From this recapitulation one may conclude that conventional computers are strong in the compound real-time control problems. These are the problems that involve several elementary calculations. The central computer in a modern weapon system faces such a problem. The capacities of the conventional computer are even more necessary when the problem includes general information processing.

Designability is an extremely important asset in these compound problems. The large, complicated systems change rapidly as they are developed and tested. The computer which can be adapted to these changes easily and quickly is much to be preferred. Modern

Table I. Comparison of Typical Computers in Air-Borne Real-Time Control

	Type of Computer		
Characteristics	Simple Incremental Computer, DDA*	Incremental Computer with Fast Slewing: Increments 0, ±1, ±32†	Conventional Computer, GP‡
Bandwidth, cps	25	25	2.5
Slewing time, sec	300	10	0.05
Fundamental speed limitation	Rate-limited, 1 part in 300/sec	Rate-limited, 1 part in 10/sec	Frequency-limited, 2.5 cps
Precision	1 part in 1,000	1 part in 1,000	1 part in 1,000
Relative volume and weight	1/2	3/4	1
		3/4	1
Relative number of parts	1/4	1½	1
Maintainability, hours down per failure	1/2		
Malfunctions per 1,000 hours			
Total	10	40	50
Serious	5	7	2
Failures per 1,000 hours	1	4	5
Reliability in unsatisfactory performances per 1,000 hours	6	11	7

The numbers are estimates, but should be correct to within a factor of 3.

* Suitable problem for simple incremental computer: Guidance of single-stage missile; a small, single problem requiring light weight but no fast slewing.

† Suitable problem for incremental computer with fast slewing (increments 0, ±1, ±32): Defensive fire control; a moderate-sized problem requiring large bandwidth and fast slewing.

‡ Suitable problem for conventional computer (GP): Central computer in bomber; a large, compound problem including decisions, data-processing, and fast slewing.

systems are built hurriedly, and there is no time to spare for redoing computers.

In the big problems, the chief handicap of the conventional computer is its modest computation frequency. When the work accumulates, the machine may not recycle fast enough for good dynamic system response.

The better reliability of conventional computers is a mixed blessing. It is composed of good reliability with respect to malfunctions and poor reliability with respect to failures. The weakness with respect to failures may be a significant handicap. However, maintainability is good.

In simpler problems, the conventional design is less desirable. Its strongest virtues are unnecessary, and its complexity is a drawback. However, where slewing time must be short, its appropriateness is indicated.

Table I summarizes the comparisons given. The numbers are estimates only. It is to be expected that actual values will be up to three times larger or smaller. They should not, however, be off by factors much larger than 3.

The Mainstream of Computer Development

The computer industry is working almost exclusively with conventional computers and will continue to do so. This is to be pondered in weighing relative merits. The industry is a large one; it produces inventions and designs, component parts, engineers and maintenance men, programs, techniques, and a literature. Much of this output cannot be used in an incremental computer without extensive modifications. An incremental project will therefore have to proceed alone to a great extent; it will find itself out of the mainstream of computer work. This argues strongly for the conventional machine.

Critical Discussion

A very real effort has been made to avoid bias in forming the opinions in this article and, though no change is apparent, a possible major change in circuit technique might of course, affect the conclusions.

A hybrid machine, part conventional and part incremental, is a tempting possibility. The conventional portion would do the general information-processing and deciding; the incremental would do the repetitive calculation upon continuous quantities. On starting up, detecting an error, or making a decision requiring a change in the course of the calculation, the conventional would provide the incremental with a new solution. Therefore the hybrid would have both short slewing time and high computation frequency. At least one such machine has been built, and is reported to be successful.

Again it should be stated that this discussion has been limited to the air-borne problems. A similar analysis for industrial process-control would be timely and it is hoped that this paper may be helpful in such an analysis.

Forecast

Considering the trend to more and more complicated systems with more and more functions done automatically, the severe limitations on the simple incremental, the small advantage gained at much cost in the incremental with multiple incremental sizes, and the course of the computer industry, it is felt that the place of the incremental computer in air-borne control is limited. It seems predictable that the conventional computer will be very generally used. Incremental techniques may be used where they can provide functions which are unnatural to the conventional computer.

References

1. DDA—WHAT IS IT?, R. N. Goldman. *Electrical Engineering*, vol. 77, no. 7, July 1958, pp. 592–94.

2. ERROR DETECTION AND ERROR CORRECTION IN REAL-TIME DIGITAL COMPUTERS, Anthony Ralston. *Proceedings*, Institute of Radio Engineers, New York, N.Y., 1957, pp. 179–88.

The Existence of an Electric Analog of a Magnetic Circuit Established by the Basic Laws Defining Magnetostatics

ELDO C. KOENIG
ASSOCIATE MEMBER AIEE

EQUIVALENT circuits have been established by Kron and others who used finite-difference and integral forms of Maxwell's equations and equations of mathematical physics in general, and in many cases applied them to actual problems in obtaining solutions. Examples of application are to problems of compressible and incompressible fluid, electromagnetic fields, elastic fields, and heat flow. See references 1 through 19.

This paper is concerned with the special case of the equivalent circuit of magnetic fields of static current circuits where flux is represented by current, and mmf (magnetomotive force) by emf (electromotive force). The equivalent has been applied in analyzing magnetic circuits containing saturable material. The analog computer used has previously been described;[16,17] it contained nonlinear resistance elements whose nonlinearity gave an ampere-versus-voltage curve of the same shape as the saturation curve of steel. This type of equivalent circuit, exemplified in Fig. 1, is to be compared with a second type.

The second type of equivalent circuit, the one most generally used,[9,14] is shown in Fig. 2. The magnetic field which is simulated in both the first and second types, is that produced by equal currents flowing in opposite directions in two conductors normal to the *xy* plane; see Fig. 3. The medium surrounding the conductors extends to infinity but only the network in the vicinity of the conductors is shown. Here, known currents are injected into or taken from the network at the junctions, and flux densities are represented by the difference of potentials between the junctions.

The first type of circuit, with which this paper is concerned, is believed to be of greater significance than the second, and deserving of added consideration for several reasons:

Paper 56-118, recommended by the AIEE Basic Sciences Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., January 30–February 3, 1956. Manuscript submitted March 28, 1955; made available for printing February 18, 1960.

ELDO C. KOENIG is with the Allis-Chalmers Manufacturing Company, Milwaukee, Wis.

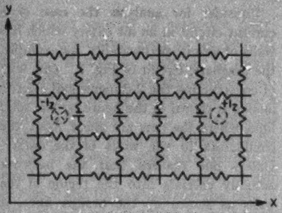


Fig. 1. Equivalent circuit with lumped resistances for magnetostatic field of Fig. 3 with battery in each circuit element at location of branch cut

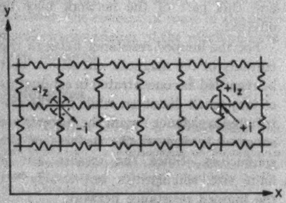


Fig. 2. Second type of equivalent circuit for magnetostatic field of Fig. 2 with currents injected into and taken from network at location of magnetic circuit conductors

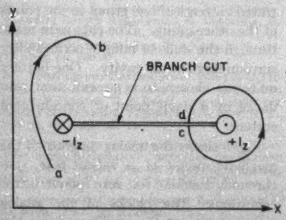


Fig. 3. Simple 2-dimensional magnetostatic field with branch cut between conductors

1. When the same basic laws are applied to magnetostatics and to current density flow, the results reveal directly the mmf-emf and flux-current correspondence.

2. The concept of the mathematical branch cut used in classical solutions is clearly perceptible in the equivalent circuit and serves as a guide in reducing the circuit to the simplest form.

3. When sketches of lines of equimagnetic potential rather than tubes of flux are made of magnetic circuits (such as those of electric machines), armature reaction, slot leakage, etc., are more easily visualized.[19] The equivalent circuit with the mmf-emf and flux-current correspondence may be easily constructed from such a sketch, and the significance of the terms related to the magnetic circuit is retained in the analog.

Establishing the Equivalent Circuit

Consider for analysis the simple 2-dimensional magnetostatic field, as shown in Fig. 3, produced only by the currents indicated which are of equal magnitude and flowing in opposite directions in two conductors normal to the xy plane. Throughout the region outside the conductors a scalar potential exists. The scalar potential U is made a single valued function of position by inserting a branch cut between the two conductors, and thus a Riemann sheet is established. Two parallel lines are shown in Fig. 3 as representing the two edges of the cut. The cut prevents the complete encirclement of the current in the calculation of the line integral:

$$\oint_a^b \mathbf{H}\cdot d\mathbf{s} = \oint_a^b \frac{1}{\mu}\mathbf{B}\cdot d\mathbf{s} = U_a - U_b \qquad (1)$$

where a and b are any two points in the open region bounding the two conductors and the cut. For the special case of a and b located on opposite edges of the cut, such as c and d in Fig. 3, the magnetic potential differences by Stoke's theorem equals the current flow in the circuit, or

$$U_c - U_d = I_z$$

To establish the analog consider a sheet of material of conductivity σ in the xy plane, as shown in Fig. 4. Circular holes are cut in this sheet at the location of the conductor of Fig. 2 and of the same diameter. A cut is made between the two conductors following a line of equimagnetic potential and the edges of the cut are plated with a material of infinite conductivity. The terminals of a battery are connected to the plated edges. Then, by the density flow equation for steady current,

$$\oint_a^b \frac{1}{\sigma} i\cdot d\mathbf{s} = V_a - V_b \qquad (2)$$

By comparison of equations 1 and 2 it is noted that magnetic potential U corresponds to electric potential V magnetic induction $\mathbf{B}$ to current density i, permeability μ to conductivity σ, and flux Φ to current I.

The same analysis applies to three dimensions where the path of the integral is a space curve. A magnetic field produced by current flowing in a single-turn conductor is shown in Fig. 5. The periphery of the cut follows the contour of the conductor. For the equivalent circuit the surfaces of the cut are coated with material of infinite conductivity and the terminals of the battery are connected to the two surfaces.

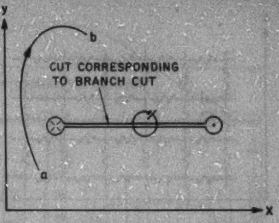


Fig. 4. Equivalent circuit of magnetostatic field of Fig. 3 consisting of sheet of conducting material with battery connected across cut

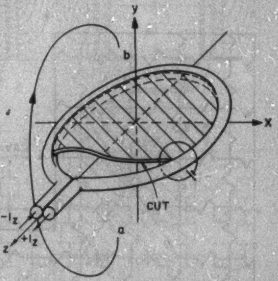


Fig. 5. Three-dimensional magnetostatic field with cut in plane of looped conductor

Equivalent Circuit with Lumped Resistance

For general application it is desirable to use a lumped resistance circuit rather than a sheet of conducting material representing distributed resistance. Fig. 1 shows an equivalent circuit of lumped resistance for the magnetic circuit of Fig. 3. The cut representing the branch cut is made along a line of equimagnetic potential in the same manner as before for the distributed resistance circuit of Fig. 4 and the ends of the resistance elements on either side of the cut connected to the terminals of batteries. It is evident that the sum of the voltage drops of any chosen mesh is equal to zero except when the mesh includes a conductor of the simulated magnetic circuit in which case the sum of the voltage drops equal the battery voltage. This is true, regardless of the manner in which the cut is made between conductors. Fig. 6 shows another way of making the cut for the same magnetic circuit.

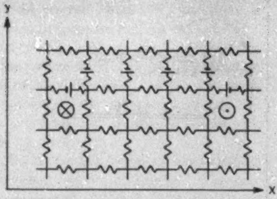


Fig. 6. Second arrangement of batteries for equivalent circuit of Fig. 1, where simulated branch cut does not follow line of equimagnetic potential

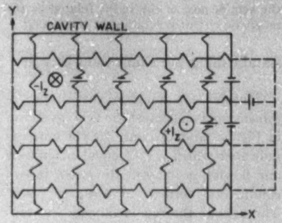


Fig. 7. Electric analog of magnetic circuit for conditions of static current circuit in air cavity with walls of infinite permeability

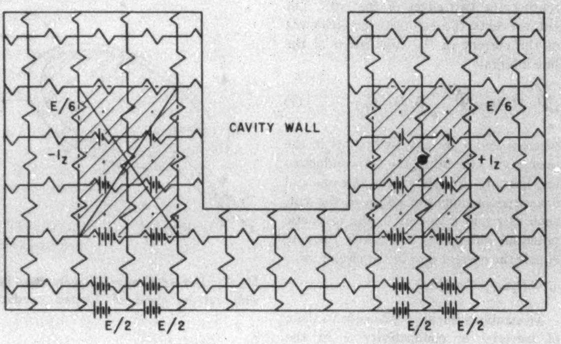


Fig. 8. Electric analog of magnetic circuit for conditions of coil, large relative to mesh size

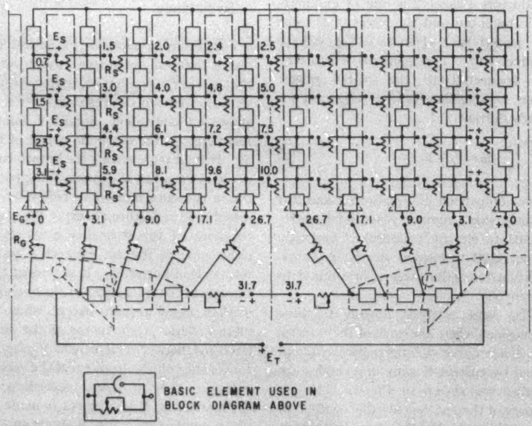


Fig. 9. Analog diagram of magnetic circuit of salient-pole synchronous machine for zero power factor conditions assuming that only teeth and pole head saturate

Consider for analysis the case of a current circuit in an air cavity which has walls of infinite permeability, for example, the magnetic circuit of Fig. 7. One way of constructing the 2-dimensional electric analog is shown in the figure by the solid lines where the cuts are made from the conductors to the walls of the cavity. Unless one realizes that the network actually extends into the region of the walls as shown by the broken lines, it would appear that the branch cut as suggested by the location of the emf was discontinuous at the edge of the cavity. Because the permeability of the material of the cavity walls is infinite, the mesh elements have zero resistance and this part of the network may be omitted.

For the lumped resistance network the currents of the magnetic circuit can best be regarded as concentrated in conductors of very small diameter, while for the distributed resistance circuit the currents are assumed to flow at the surfaces of the conductors. Since the meshes are of finite size, singularities are avoided in the lumped resistance network.

The example circuit of Fig. 8 illustrates how current in a group of conductors may be considered to be concentrated at a number of points in the center of the meshes of the analog. Cuts are made from each of these six points of concentrated currents of one group to the points of the other group. The cuts were made through the walls of infinite permeability surrounding the air cavity. One battery unit was chosen to represent mmf produced by a single point of concentrated current.

Fig. 9 shows the analog diagram of the magnetic circuit of a salient-pole synchronous machine for zero power factor conditions. The blocks of the analog circuit represent the nonlinear basic elements for simulating magnetic saturation of the stator teeth and pole tips. The cuts are made in such a manner as to agree with the sketch of lines of equimagnetic potential (ampere-turn lines) of Fig. 10, and the emf E_s and E_g are placed in the network at these cuts. The significance of such terms as armature reaction and leakage reaction remain perceptible in the analog circuit. The approximate relative magnitudes of the emf are written on the network drawing.

By sectioning the air gap with the aid of a rough sketch of the field, it is possible to reduce the number of resistance elements of the network in this region. The average dimensions of the resulting odd-shaped sections are normally taken in calculating the resistance values. When odd-shaped sections of saturable material are considered, such as the pole tip sections, it may be desirable to adjust the width from the average value according to the degree of saturation or to use a finer mesh circuit. For 2-dimensional fields the depth is normally taken as unity.

Fig. 11 shows another arrangement of emf for the analog of the magnetic circuit of Figs. 9 and 10, in which case the cuts in the network corresponding to the branch cuts are made from the slot conductors to be bottom of the slots and on through the core, which is assumed to have infinite permeability. It is seen that the number of emf have been reduced and that those eliminated are of the largest magnitude; however, the concept of some of the electrical characteristics of the machine have been obscured.

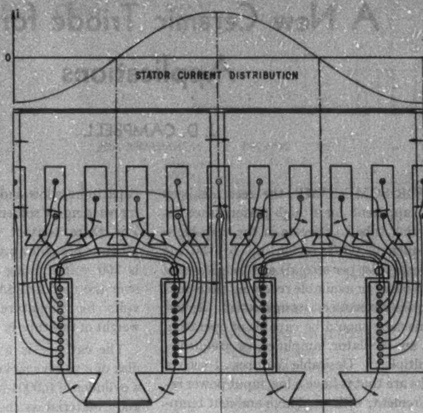


Fig. 10. Sketch of pair of poles and portion of stator of salient-pole synchronous machine showing lines of equimagnetic potential (ampere-turn lines) for zero power factor conditions to serve as aid in locating emf in analog of Fig. 9

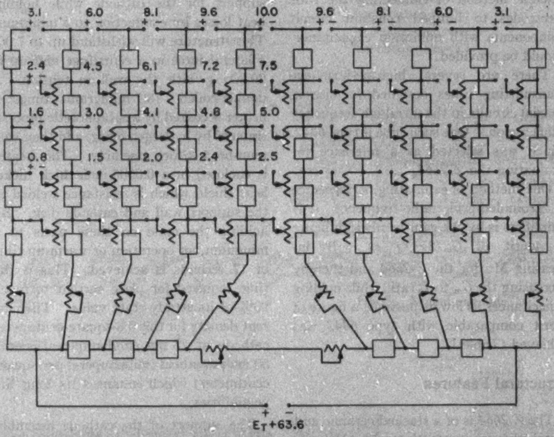


Fig. 11. Analog block diagram for salient-pole machine of Figs. 9 and 10 showing different choice of locations of emf determined by assuming that branch cuts of magnetic circuit extend to bottom of stator slots

References

1. A Critical Study of Variational and Finite-Difference Methods for Calculating the Operating Characteristics of Waveguides and Other Electromagnetic Devices, R. M. Soria, T. J. Higgins. National Electronics Conference, Chicago, Ill., Nov. 1947.

2. Equivalent Circuits of the Electromagnetic Field, U. F. McAllister, Jr. *General Electric Review*, Schenectady, N. Y., vol. 47, 1944, pp. 9–14.

3. Numerical Solutions of Ordinary and Partial Differential Equations by Means of Equivalent Circuits, G. Kron. *Journal of Applied Physics*, New York, N. Y., vol. 16, no. 3, 1945, pp. 172–86.

4. Network Analyzer Studies of Electro magnetic Cavity Resonators, C. Concordia, G. Kron, W. Ridgeway, J. R. Whinnery. *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 32, 1944, pp. 360–67.

5. A Method for Determining Unsteady-State Heat Transfer by Means of an Electrical Mode, V. Paschkis, H. D. Baker. *Journal*, American Society of Mechanical Engineers, New York, N. Y., vol. 64, 1942, pp. 105–12.

6. Numerical and Network-Analyzer Solution of the Equivalent Circuits of the Elastic Field, G. K. Carter. *Journal of Applied Mechanics*, New York, N. Y., vol. 11, 1944, pp. 162–67.

7. Equivalent Circuits of the Elastic Field, G. Kron. *Ibid.*, pp. A149–61.

8. Equivalent Circuit of the Field Equation of Maxwell, G. Kron. *Proceedings*, Institute of Radio Engineers, vol. 32, 1944, pp. 289–99.

9. Electric Circuit Models of Partial Differential Equations, G. Kron. *Electrical Engineering*, vol. 67, no. 7, July 1948, pp. 672–84.

10. A String-Net Analog for the Numerical Solution of the Equations of Laplace and Poisson, Peter L. Tea. *Journal*, Franklin Institute, Philadelphia, Pa., vol. 258, 1954, pp. 287–303.

11. A New Approach to the Solution of High-Frequency Field Problems, J. R. Whinnery, S. Ramo. *Proceedings*, Institute of Radio Engineers, vol. 32, 1944, pp. 284–89.

12. Equivalent Circuits to Represent the Electromagnetic Field Equations, G. Kron. *Physical Review*, New York, N. Y., vol. 64, 1943, pp. 126–28.

13. Electric Circuit Models of the Schrödinger Equation, G. Kron. *Ibid.*, vol. 67, 1945, pp. 39–43.

14. Equivalent Circuits of Compressible and Incompressible Fluid Flow Fields, G. Kron. *Journal of the Aeronautical Sciences*, New York, N. Y., vol. 12, no. 2, 1945, pp. 221–31.

15. Numerical and Network-Analyzer Tests of an Equivalent Circuit for Compressible Fluid Flow, G. K. Carter, G. Kron. *Ibid.*, pp. 231–43.

16. An Electric Analogue Computer Using the Photo Cell as a Nonlinear Element, Eldo C. Koenig. *AIEE Transactions*, vol. 69, pt. II, 1950, pp. 1606–11.

17. Photoelectric Analog Computer, Eldo C, Koenig. *Electronics*, New Yok, N. Y., Dec. 1950, pp. 124–26.

18. Electromagnetic Theory (book), J. A. Stratton. McGraw-Hill Book Company, Inc., New York, N. Y., 1941.

19. Approximating Potier Reactance, Sterling Beckwith. *AIEE Transactions* (*Electrical Engineering*), vol. 56, July 1937, pp. 813–18.

A New Ceramic Triode for VHF Applications

J. D. CAMPBELL
NONMEMBER AIEE

FOR SOME TIME, the need has been apparent for a high-performance receiving tube to operate in the very-high-frequency (vhf) region up to 450 mc (megacycles per second) and to bridge the gap between available receiving tubes and high-performance transmitting tubes. This tube should be capable of operating as an oscillator, amplifier, or frequency multiplier. Desirable features of such a tube are that it have a low input power requirement, withstand high ambient temperatures, and be as small as possible. A reliable method of connecting the tube electrodes to lumped constant circuit components with minimum capacitance, should be provided.

There are several high-performance transmitting tubes designed for use in coaxial circuits in the ultrahigh-frequency (uhf) region. One such tube is type *6442*, which was selected as a reference for evaluating the *Z-2664*.

One method of estimating performance of grounded-grid radio-frequency (r-f) amplifiers is by the gain-bandwidth figure of merit of $G_m/2\pi\sqrt{C_{in}C_{out}}$. By increasing Mu for the *Z-2664* and thereby increasing the G_m-to-I_p ratio while holding capacitances as low as possible, a figure of merit comparable with type *6442* was achieved (Table I).

Structural Features

The *Z-2664* is of a stacked ceramic and metal planer construction with electrode contacts protruding from the side of the cylindrical assembly to allow socketing, soldering, or welding to associated circuitry (Fig. 1). All contacts are spaced at multiples of 0.100 inch to conform to standard printed circuit modules. The outside diameter of 0.550 inch is determined by the minimum allowable cathode area while the longitudinal spacing between electrode contacts is set at the optimum multiple of 0.100 inch to keep capacitances low and provide ample room for welding or soldering to the terminals. For these contact spacings, the high-altitude breakdown voltage at 80,000 feet is 700 volts, giving ample safety factor over the new ASESA requirement of 500 volts for miniature tubes. The total weight of this tube is 10 grams.

The cathode is a 0.260-inch-diameter disk of a passive nickel alloy supported on a cylinder of 0.0004-inch-thick nickel-iron alloy material as shown in Fig. 2. The thin wall of the cylinder provides a rigid support for the cathode while holding heat losses by conduction to a minimum. This structure will withstand up to 7.000 *g* lateral force in a centrifuge and up to 20,000 *g* with the wall in tension. A double-coiled 13.7-milligram tungsten heater is coated by a cataphoretic process, sintered by hydrogen firing, and then vacuum-fired before assembly. This heater is enclosed in a 0.0005-inch-thick nickel heat shield which is resistance welded to the support wall and cathode disk. By holding the mass on these parts to a minimum, an operation or warm-up time of 17 seconds is achieved. This is the time required for plate current to reach 90% of its steady-state value. The current density for this 0.3-square-centimeter cathode at 15 ma (milliamperes) is only 50 ma per sq cm (milliamperes per square centimeter) which enhances its long life possibilities.

The support of the cathode assembly is spot welded to a titanium ring with the distance between the top surface of the ring and the top of the cathode held to a tolerance of 0.0005 inch. The cathode is sprayed with a carbonate coating which is shaved to give a coated height 0.003 inch higher than the metal cover.

The size of the cathode requires that the inside diameter of the grid support be 0.300 inch to give adequate clearance. Since a parallel wire frame grid would allow excessive vibration output due to individual wire resonances, a woven mesh is used (Fig. 3). A disk is cut from 0.001-inch-diameter molybdenum wire cloth, having a nominal thickness of 0.002 inch and pressed to give an over-all thickness of 0.0014 inch. This pressing serves to lock the wires in position and place all portions of the wire mesh in essentially the same plane, which is desirable for a uniform grid characteristic. The mesh, a copper brazing shim, and a tungsten support washer are placed, in that order, in a 0.001-inch-thick molybdenum support cup. This assembly is stacked between ceramic spacers and fired in a hydrogen atmosphere slightly above the melting point of copper. The support cup of the brazed assembly is then formed flush with the top of the tungsten support. It will be noted that the side of this cup is sloped to allow for the difference in thermal expansion of the tungsten washer and the titanium ring on which this cup will be supported.

The vibration output from the tube using this grid is shown in Fig. 4. The natural reasonant frequency at which peak output is obtained is 10,000 cps (cycles per second), with the smaller peak outputs occurring at subharmonics of the frequency. However, these peaks are no higher than the output from a typical receiving tube such as the *5670*, which has only one half the gain of the *Z-2664*. The integrated level of noise on the *Z-2664* would be much less than that for the *5670*.

The grid assembly is supported by a titanium ring similar to that used for the cathode and held in place by a ceramic spacer which is ground to a thickness tolerance of ±0.0005 inch (Fig. 5). This spacer, the 0.010-inch-thick tungsten grid support, and the shaved cathode height combine to give a cold grid-to-cathode spacing of 0.0035 inch. When the cathode is heated, the cathode support expands to produce a spacing of 0.003 inch. This spacing results in a transit time too large for uhf applications but one that is quite satisfactory for vhf, with particular advantages gained in low input capacitance. Greater freedom from short circuits and reduced spread of characteristics is possible with this conservative spacing.

The anode is made by welding a 0.265-inch-diameter titanium stud to the titan-

Table I. Gain-Bandwidth Figure of Merit

$$\text{Figure of Merit} = \frac{G_m}{2\pi\sqrt{C_{in}C_{out}}}$$

	Type 6442	Type Z-2664
I_p	35 ma (max.)	15 ma
G_M	16,500 micromhos	15,000 micromhos
M_μ	50	80
C_{in}	5.1 μμf	4.6 μμf
C_{out}	2.3 μμf	2.0 μμf
Figure of Merit	700	770

Paper 60-75, recommended by the AIEE Electronics Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., January 31–February 5, 1960. Manuscript submitted July 28, 1959; made available for printing March 1, 1960.

J. D. Campbell is with the General Electric Company, Owensboro, Ky.

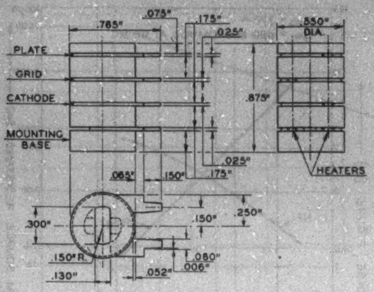

.765"
.075"
.175"
.025"
.350" DIA.
PLATE
GRID
CATHODE
MOUNTING BASE
.875"
.025"
.175"
.065"
.150"
HEATERS
.300"
.150"
.250"
.150"R.
.130"
.052"
.080"
.006"

Fig. 1. Outline

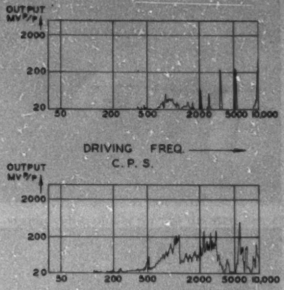


Fig. 4. Swept frequency vibration test

Top—Z-2664 test conditions:
$R_L = 10$ kilohms
Acceleration = 10 g
10 minutes per sweep
Test circuit gain = 52
Bottom—5670 test conditions:
$R_L = 10$ kilohms
Acceleration = 10 g
10 minutes per sweep
Test circuit gain = 21

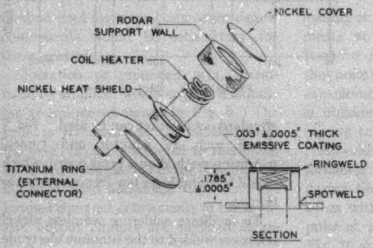

RODAR SUPPORT WALL
NICKEL COVER
COIL HEATER
NICKEL HEAT SHIELD
.003" ±.0005" THICK EMISSIVE COATING
RINGWELD
TITANIUM RING (EXTERNAL CONNECTOR)
.1785" ±.0005"
SPOTWELD
SECTION

Fig. 2. Cathode assembly

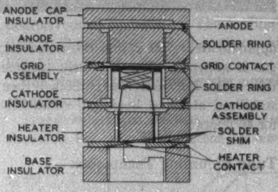

ANODE CAP INSULATOR
ANODE INSULATOR
GRID ASSEMBLY
CATHODE INSULATOR
HEATER INSULATOR
BASE INSULATOR
ANODE
SOLDER RING
GRID CONTACT
SOLDER RING
CATHODE ASSEMBLY
SOLDER SHIM
HEATER CONTACT

Fig. 5. Mount assembly

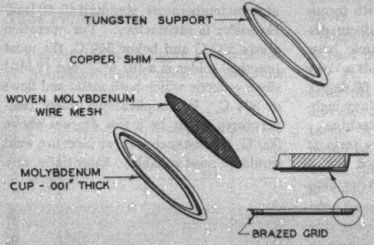

TUNGSTEN SUPPORT
COPPER SHIM
WOVEN MOLYBDENUM WIRE MESH
MOLYBDENUM CUP - .001" THICK
BRAZED GRID

Fig. 3. Grid assembly

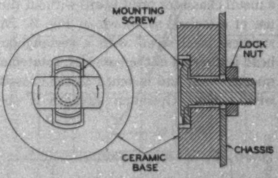

MOUNTING SCREW
LOCK NUT
CHASSIS
CERAMIC BASE

Fig. 6. Application of mounting base

ium anode contact. This assembly is then formed to a study height tolerance of ±0.0005 inch. With a grid to anode spacing of 0.018 inch, characteristics are not appreciably affected by minor dimension changes, and a relatively low value of grid to plate capacitance is obtained.

The thin ceramic cap at the top of the anode is used primarily to add strength to the anode seal. This additional strength is needed during thermal shock caused by welding or soldering components to the anode contact or when rated dissipation is accidently exceeded by a large amount. A plate dissipation of 12 watts, which is four times the tube rating, will not damage the seal. However, this is definitely not recommended as an operating condition.

All parts of this tube are vacuum-fired prior to assembly and nickel solder shims are spot welded to the perimeter of the titanium contact parts on both top and bottom sides. The components are stacked in jigs which align the electrodes, and axial pressure is applied to each end of the tube. Loaded jigs are placed in a vacuum furnace and pumped to less than 0.1 micron of mercurial pressure. Heat is applied until the cathode has been activated, with gases escaping through the unsealed joints. The temperature is then increased to slightly above 970 C (degrees centigrade), the temperature at which titanium and nickel form a eutectic and the tube is sealed. Since this process uses a portion of the titanium metal adjacent to the nickel shim for solder, the shim is designed so that it does not cover the entire support area between the insulator and contact ring; otherwise, spacings would shift upon sealing.

Even with the high vacuum in the furnace at the time of sealing, a small amount of gas remains in the tube after sealing and some additional gas is released during initial aging. The extremely good gettering properties of titanium at elevated temperatures are utilized during aging. Heater voltage is applied to the tube in a vacuum furnace at a temperature of

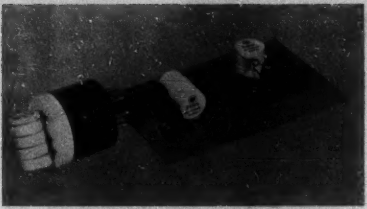


Fig. 7 (left). Low-frequency mounting methods

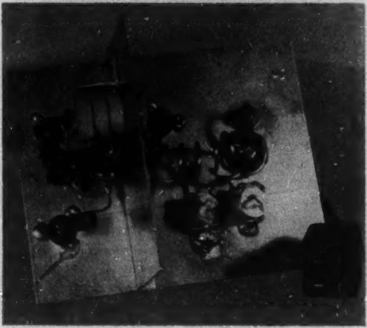


Fig. 8 (left). High-frequency mounting technique

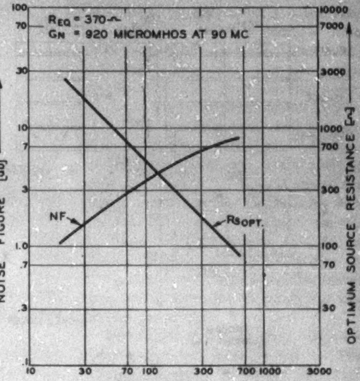


Fig. 9 (right). Predicted noise performance

approximately 750 C and the remaining gas is gettered. The relatively gas-free tube with a fully activated cathode is then stabilized at rated operating conditions for 46 hours.

Tube Mounting Techniques

The ceramic mounting base, which is sealed to the heater end of the tube, allows the tube to be fastened directly to a metal chassis or print board without the use of a socket, as shown in Fig. 6. A stainless steel screw with a rectangular head is inserted in the base slot, rotated a quarter-turn, and fastened to the chassis with a nut. Optimum strength is obtained by tightening the nut with torque of 3 inch-pounds, giving a uniform distribution of force on the ceramic base. This type of mounting withstands a 720 g shock which is equivalent to a 60-degree hammer angle on the standard Navy high-impact (flyweight) shock test machine.

For performing low-frequency tests at room temperature, an adaptor for a standard octal socket can be used. Fig. 7 shows such an adaptor and typical mounting method on a print board, which would be suitable for low-frequency operation.

To facilitate soldering seamless nickel boots are welded to the titanium electrode contacts. With circuits to be operated at room temperature, standard 40–60 lead-tin solder is satisfactory. For operation above 200 C and below 300 C, the most practical solder is a (97.5%) lead (2.5%) silver eutectic which both melts and flows at 304 C. For applications where temperatures might be a few degrees above 300 C, a resistance weld or inert arc weld would be most reliable. Essentially, any

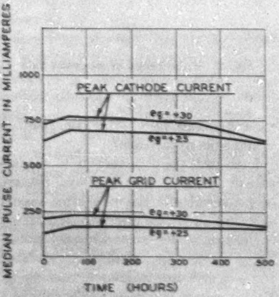


Fig. 10. Pulse life test

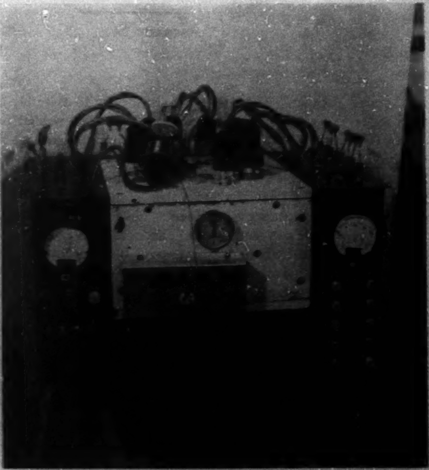


Fig. 11. High-temperature life test oven

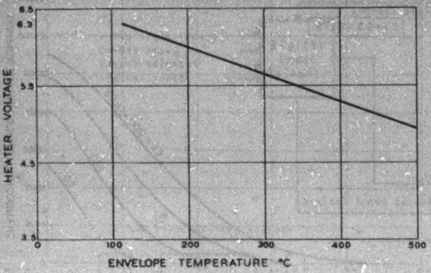


Fig. 12. **Heater derating for constant pulse emission**

method used for connecting other circuit components would be satisfactory for these contacts.

Operation as grounded grid, broadband r-f amplifier would suggest a mounting technique as shown in Fig. 8. The tube is mounted on the shield plate; the grid contact is soldered to this plate, and a bracket attached to the tube mounting base gives additional support. The lumped constant components of the input and output circuits are shielded as required.

Tube Performance

A calculated noise figure as a function of frequency, based upon transit time loading and shot noise resistance measurements at 90 mc, is plotted in Fig. 9. Because of the relatively low values of shot noise resistance, lead inductance, and capacitances, a better effective noise figure is obtained than with conventional receiving tubes.

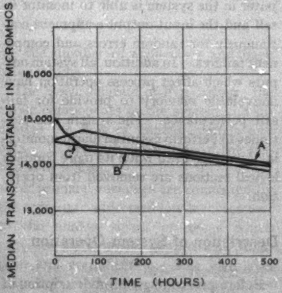


Fig. 13. **Transconductance versus life**

Test conditions:
$E_f = 6.3$
$E_p = 200$ volts
$R_k = 68$ ohms
Life test conditions:
$E_p = 250$ volts
$R_k = 130$ ohms
$W_p = 3.3$ watts
A—$E_f = 6.3$
Room ambient temperature
B—$E_f = 6.3$
300 C envelope temperature
C—$E_f = 5.65$
300 C envelope temperature

An evaluation of the stability of r-f performance in class *B* and *C* applications during life is made by pulse-life testing tubes with the standard 10-microsecond pulse at a 1,000 pulses per second rate. Fig. 10 shows the peak cathode and grid currents during life for two different grid drives.

High-temperature life tests are performed in ovens such as that shown in Fig. 11. Lead wires from the octal plug enter the oven through a ceramic shield and are resistance welded to the tube contacts. A thermocouple cemented to the center of the anode cap insulator is used to set the oven ambient temperature to give 300 C envelope temperature. There is no forced circulation in the oven so that this ambient is normally 240 C with the remaining heat created by tube dissipation. These portable ovens are rolled periodically to a test set and tube readings are made while the oven is operating at elevated temperatures to obtain a true indication of tube performance.

To secure the longest possible life from this tube when operated in a high ambient temperature, the heater voltage should be reduced by an amount necessary to prevent any rise above normal cathode operating temperature. Pulse emission is one of the most sensitive methods for measuring changes in this temperature. The heater derating curve in Fig. 12 shows the heater voltage necessary to maintain constant pulse emission for increasing envelope temperature. At the maximum rating of 300 C for this tube, the heater should be operated at 5.65 volts.

Although transformers can be purchased which will track this derating curve, there is some reluctance on the part of equipment designers to make this special provision. Therefore, this tube is now rated for 300 C envelope temperature operation at 6.3 volts without the necessity of heater derating if extremely long life is not a consideration.

Fig. 13 shows the life test performance under derated and nonderated conditions as well as room ambient operation at rated heater voltage. Although the difference at 500 hours is not extreme, a more stable value of *Gm* for the derated heater condition can be expected for longer periods of life.

Since in practice, most equipment will be adjusted at room ambient temperature and then expected to operate at elevated temperature, it is desirable to know what shift in characteristics to expect. Pulse emission, being extremely sensitive to cathode temperature, increases by approximately 50% when the tube is changed from room temperature environment to 300 C envelope temperature operation. Under the same conditions, plate current and *Gm* will increase by less than 3%.

Possible early life changes are controlled by reading *Gm* after 1 hour and after 20 hours of operation under class *A* conditions at room temperature. The distribution of *Gm* at these periods is shown in Fig. 14.

The plate current characteristic curves are shown in Fig. 15.

Transconductance transfer curves are shown in Fig. 16. Here, it can be seen

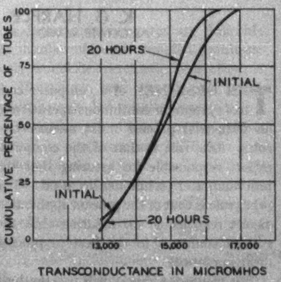


Fig. 14. **Distribution of transconductance on stability life**

Test conditions:
$E_f = 6.3$
$E_p = 200$ volts
$R_k = 68$ ohms
Stability life
test conditions:
$E_f = 6.3$
$E_p = 250$ volts
$R_k = 130$ ohms
Room ambient temperature
Sample Size
25 tubes

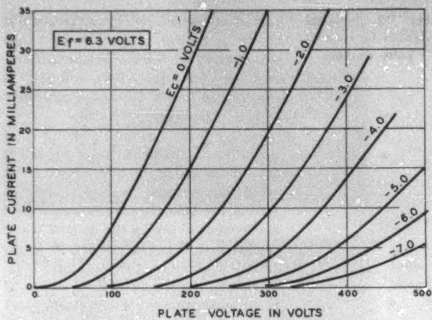


Fig. 15. **Average plate characteristics**

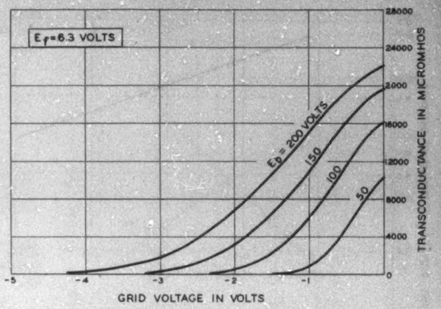


Fig. 16. **Average transfer characteristics**

that for zero bias operation, the rated *Gm* of 15,000 can be obtained with approximately 100 volts on the plate.

Conclusions

The *Z-2664* as a vhf tube has a low cost compared with other tubes with equivalent performance. Its low power requirement and relatively small size make it especially attractive for mobile or airborne equipment use. Microphonic output is less than that normally obtained with glass receiving tubes. Elimination of socketing further reduces space requirements and eliminates oxidized contact problems at high temperatures. The maximum temperature rating of 300 C is as high as the rating on most circuit components now available. With the bulk of the tube parts composed of ceramic and titanium, deterioration due to atomic radiation is very slight. The geometry of its structure makes it more resistant to shock and vibration than any conventional assembly glass tube which might do the same job.

Vol.
Pt. I
196

The LN 3000 Computer Control System

K. G. HARPLE
ASSOCIATE MEMBER AIEE

R. G. LEX, JR.
ASSOCIATE MEMBER AIEE

THE DESIGNER of a computer control system for continuous operation in an industrial process must assume two roles. One role is that of the computer expert responsible for assuring that the best computer techniques are used. The other role is that of the process application expert responsible for assuring that the system conforms to the best in process design practices.

The computer expert will use the best in components and circuit design to meet the stringent reliability and performance requirements. He will employ solid-state components, circuits will be conservatively rated, and particular attention will be given to manufacturing techniques. The computer expert feels that he now has equipment acceptable for continuous operation in an industrial process.

While the process expert may agree that the equipment will exhibit excellent reliability, he will quickly ask what error checking and fail-safe performance techniques have been used. He will point out that even the most reliable process equipment employs interlocks to provide fail-safe operation. On this basis, the process expert will insist that these features be incorporated in the computer control system before it is installed on a process.

The system described in this paper was designed using both points of view. The circuits in the system were exhaustively designed. Error detection and correction were incorporated through programming and logical design techniques. The computer in the system is able to monitor itself and the input-output equipment continuously for random errors and component failures. In addition, all system outputs which affect process operation have nonvolatile memory to provide for fail-safe performance. The system also continues to perform essential alarm monitoring functions even when its more complex logical sections are removed from operation.

Description of System Operation

Before proceeding with a description of the system operation and design features, a few brief remarks on the circuits may provide a helpful background. The logic circuits are all built around one basic diode-transistor module. The wide use of a thoroughly designed single circuit is advantageous from the reliability and maintenance standpoint. Since amplification provided by the transistor after each logical operation, a high degree of logic signal standardization is maintained throughout the equipment. Finally the design lends itself to a piggy-back module chassis card construction thereby reduc-

---

Paper 60-215, recommended by the AIEE Computing Devices Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., January 31–February 5, 1960. Manuscript submitted November 30, 1959; made available for printing March 8, 1960.

K. G. Harple and R. G. Lex, Jr. are with the Leeds and Northrup Company, Philadelphia, Pa.

This paper is largely a report of the work of many individuals, to each of whom the authors are personally indebted. Particular attention is drawn to the fact that the computer was designed and is being produced by the Computer Division, Philco Corporation.

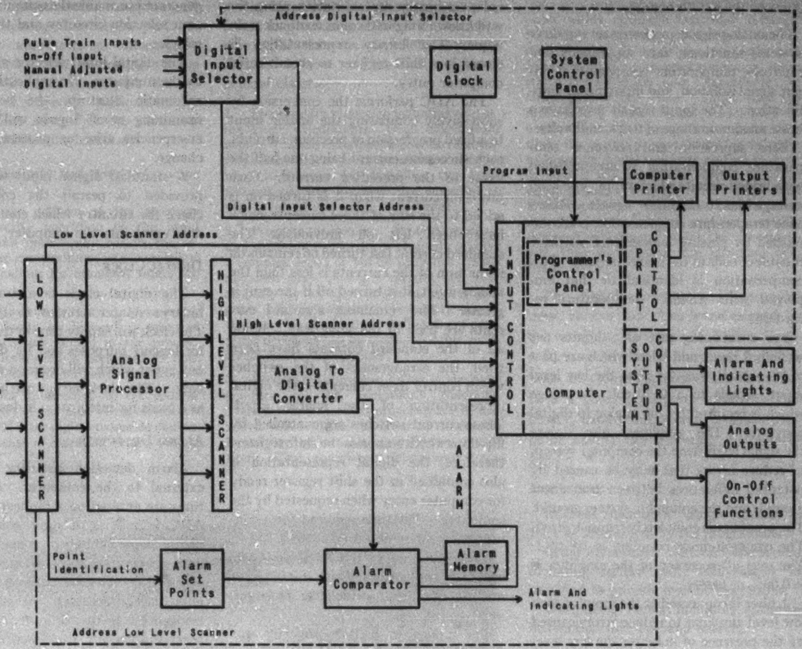

Address Digital Input Selector
Pulse Train Inputs
On-Off Input
Manual Adjusted
Digital Inputs
Digital Input Selector
Digital Clock
System Control Panel
Program Input
Computer Printer
Output Printers
Digital Input Selector Address
Low Level Scanner Address
Low Level Process Transducers
LOW LEVEL SCANNER
Analog Signal Processor
HIGH LEVEL SCANNER
High Level Scanner Address
Analog To Digital Converter
INPUT CONTROL
Programmer's Control Panel
PRINT CONTROL
SYSTEM OUTPUT CONTROL
Computer
Alarm And Indicating Lights
Analog Outputs
On-Off Control Functions
ALARM
Point Identification
Alarm Set Points
Alarm Comparator
Alarm Memory
Alarm And Indicating Lights
Address Low Level Scanner

Fig. 1. **Block diagram of the LN series 3000 computer control system**

ing the number of connectors and back panel wiring.

The block diagram of Fig. 1 shows the major components of the system and indicates the flow of control signals and data. In brief, the operation of the system is as follows.

All primary variables are connected by the computer-addressed low-level scanner to the analog signal processor which provides input signal suppression, isolation, and amplification. The high-level output from the signal processor is sequenced to an analog to digital converter by a high-level scanner.

Digital data from the analog to digital converter and a digital address corresponding to the data are transmitted to the computer input section for entry into the computer.

All digital inputs are connected to the computer input section by a computer-addressed digital input selector. Digital time from the clock is continuously available at the computer input section.

All analog and digital data are fed into the computer as demanded by the computer program. Preprogrammed calculations are performed by the computer on all data. The results of the calculations can be stored in memory for later use or can be used to make decisions for immediate control action.

As required, the computer can initiate scheduled or on-demand printouts on any of several available printers. The computer can initiate control actions by actuating on-off outputs to start and stop equipment or by readjusting the set point of a local analog controller.

To give a better understanding of the system, each of the major components of the block diagram will be discussed in turn with respect to system function and important system design considerations.

Input Switching

All analog process variables are switched by the low-level scanner. Normally the low-level scanner samples 5 points simultaneously at a rate of approximately one group per second; this gives a basic scanning speed of 5 points per second.

If faster scanning rates are desirable, the number of points sampled simultaneously would be increased. For example, at 10 points per second, 10 points would be sampled simultaneously.

The basic system has been designed to handle a maximum of 1,275 low-level analog inputs. Each low-level input signal is switched through a separate double-pole contact of a specially designed 12-pole mercury-wetted contact relay. The relay was designed to give outstanding low-level switching characteristics. Thermal emf's generated are on the order of 1.0 microvolt and transient noise is undetectable. The scanner selection is normally controlled by the computer. In control operation this permits the computer to change the scanner sequence as changing process conditions or the mode of process operation dictates.

A digital address, corresponding to the group of input points selected, is also generated by using contacts on the mercury-wetted relay. This address is used by the computer as a check on the scanner selection circuitry and scanning relays.

Analog Signal Processor

The analog signal processor serves three systems functions; zero suppression and reference temperature compensation, input signal isolation, and input signal amplification. The input signals may have a basic maximum range of 0–100 millivolts.

Zero supression and reference temperature compensation is accomplished with bridge circuits which supply low-level voltage bias to the input signals. Reference temperature compensation is accomplished by placing temperature sensitive resistance coils in the bridge circuit. This compensation is identical to that employed with standard temperature recorders.

For a scanning rate of 5 inputs per second, 5 input amplifiers which are part of the signal processor raise the low level input signals to a 0–10-volt d-c range which is required by the analog to digital converter. The amplifiers also isolate the signal leads from the computer system to reduce errors that may be caused by potential differences between transducer grounds and the computer system ground. The amplifier is completely transistorized. The common-mode rejection is 10^6 to 1. The over-all accuracy of the amplifier is ±0.06% or better.

A filter is connected at the input of each low level amplifier to reduce errors caused by the presence of noise on the low level input signals. The filter will attenuate a 60-cycle noise signal by 300 to 1. The time for the filter to stabilize to closer than 0.1% of its final value for a step change on an analog input is less than 1 second. After simultaneous selection of the low-level inputs, a one-second delay is used before any analog signals are processed by the analog to digital converter. As previously stated, 5 input amplifiers are required to obtain a scanning rate of 5 points per second. Ten amplifiers are required to obtain a scanning rate of 10 points per second.

High-Level Scanner

A high-level scanner switches the output of the 5 input amplifiers to the analog to digital converter. High-speed mercury-wetted contact relays (2-millisecond operate time) are employed as switching devices. An address is generated by relay contacts of the high-level scanner. This address is used by the computer to check the operation of the high-level scanner control circuitry and relays.

Analog to Digital Converter

The ADC (analog to digital converter) accepts the outputs of the input amplifiers in sequence and converts this analog signal to a binary representation using the well-known weighted-count feedback technique. The binary representation is stored in a shift register to permit serial computer entry.

The ADC performs the conversion by successively comparing the analog input to a fixed progression of precision currents, each successive current being one-half the value of the preceding current. Each precision current when it is turned on is added to the sum of those currents which have been left on previously. The standard current last turned on remains on if the sum of the currents is less than the analog input, it is turned off if the sum is greater. The remaining standard currents are tried in the same way. When all of the standard currents have been tried, the arrangement of the switches which control these currents is the digital representation of the analog input. These current switches are controlled by flip-flops which comprise the shift register; therefore, the digital representation is also contained in the shift register ready for computer entry when requested by the computer. The time required for a conversion is 1.5 milliseconds.

The resolution of the ADC is 10,000 counts. The accuracy is ±0.05% of reading or 1 volt, whichever is larger, ±1/2 count.

The analog to digital converter was designed to operate as a self-contained unit. The unit is not dependent on the computer for timing signals or power. The advantages of this feature will be discussed later.

Several standard analog inputs will be connected to each system for processing by the analog input section. The computer can run preprogrammed checks on the converted digital data to check both ADC and input amplifier operation.

Digital Input Selector

The digital input selector connects all digital inputs to the input control section of the computer. Digital inputs can consist of contact closures from the process, pulse train counter inputs or manually adjusted digital inputs.

The basic system has been designed to handle a maximum of 225 counter and manually adjusted inputs or 2,040 on-off inputs. Digital inputs can be read into the computer at a maximum rate of 480 inputs per second.

Mercury-wetted contact relays are used to select all digital inputs and provide isolation from computer system voltage supplies and grounds.

A digital address corresponding to the group of selected digital input points is generated to enable the computer to check scan selection circuitry and the selection relays.

The digital input selector is controlled by the computer. In applications such as automatic start-up, the sequence of examining on-off inputs will change as emergencies arise or process conditions change.

A standard digital input will also be provided to permit the computer to check the circuitry which controls digital data entry into the computer.

Digital Clock

The digital clock is a transistorized binary counter precise to 0.1 second. The clock will supply time to the computer for logging purposes and for determining any event which will occur on a scheduled basis. The clock output can also be used as a basis for integration if desirable.

Alarm Detection

Alarm detection circuitry is offered external to the computer. Alarm settings are entered on a pinboard. As the ADC is making a data conversion, a digital alarm comparison is made. Under normal operation if an alarm, either high or low, is detected, an annunciator light is immediately actuated and a signal is transmitted to the computer. The computer can execute an off-normal printout and take whatever action is required to correct the off-normal.

Since alarm comparisons could be programmed in the computer it is desirable to discuss the advantages of external alarm detection with respect to system design. When the computer is not in service, the operator can actuate a switch on the control panel which will disconnect the computer from the input equipment, leaving all essential input equipment with power. The low-level scanner will automatically become self-sequencing and the input circuitry will continue to cycle. The ADC will operate and alarm comparisons will be made since, as previously mentioned, the ADC is a self-contained unit with its own power supply and timing circuits.

If an alarm is detected in the self-sequencing state an annunciator light will be actuated. By use of alarm memory the light will remain actuated until the point alarmed returns to normal. It is also possible to log on a printer the point number and time at which the point was alarmed.

The Computer

The operation of the entire system is controlled by the computer. One may

Vol
Pt.
196

consider the operations the computer performs as consisting of two basic parts: 1. the routine mechanics of information handling, and 2. the processing of the information in accordance with the requirements of a specific application.

In the first category, the computer performs the following functions:

1. Selects the order and timing of reading input signals.
2. Checks the address of the incoming points to verify operation of the input switching, and periodically checks the accuracy of the measurement and conversion circuits by calling for standard input signals.
3. Transmits information to the operator either periodically, on demand, or when an abnormal situation occurs in the process. Operator information is presented in both printed log form and on panel indicators.
4. Transmits information to the process, both digitally, in the case of on-off control functions, and after digital to analog conversion, in the case of resetting analog control set points.

The functions which the computer performs in processing information for specific applications may be briefly summarized as follows. They fall in three general classes:

1. Process data reduction including operating guide calculations. Such calculations typically involve complex algebraic relationships between the process variables. Statistical techniques for handling the large amount of data involved are expected to play a role of increasing importance.
2. Program controller. Process start-up and shut-down procedures typically involve a complex series of yes-no decisions. Such decisions may be made rapidly and consistently by an appropriate computer program. Either completely automatic operation may be obtained, or the results of computer decisions can be made available as instructions to the operator.
3. Process optimization. This is the most complex of the computer's functions. Today, four general approaches are in use, which will be listed in order of decreasing knowledge of process optimization conditions. First, completely determinate equations may exist, whose solution by the computer yields optimum controller settings. Second, a completely determinate mathematical model of the process may exist and be stored in the computer's memory, upon which the computer may "experiment" in either real or fast time. Third, only the form of the mathematical model may be available, in which case the computer analyzes the actual process operation in order to keep the model complete and up to date, so that it will provide a correct basis for optimization. Finally, with very little knowledge about the process, the computer may be programmed to perform carefully designed experiments upon the process itself, thus seeking out the optimum controller settings.

The digital computer which provides these various functions was designed by the computer Division of Philco Corporation to Leeds and Northrup specifications, in a jointly sponsored development. It is a general purpose, serial, internal binary, fixed-point, stored program computer having the following characteristics.

Instructions

Address Structure: Single address or 1+1 as determined by the programmer.

Order Structure: Sixteen basic orders, including square root, which may be modified to provide over 60 operations.

Computing Times: including optimum instruction and operand search,

Add, subtract, or 11 other instructions = 0.91 millisecond

Multiply, divide, or square root = 3.6 milliseconds

Shift (variable, depending on number of places) = 0.91–3.6 milliseconds

Data

Word Length: 20 bits+sign+parity.

Storage

Main Memory: Magnetic drum, 8,064 or 16,256 words.

Rapid Access: Four words standard. From a programming point of view, these appear as random access cells.

Arithmetic Registers: Three transistorized registers including the accumulator, two of which may be used as additional rapid access storage.

Miscellaneous Features

Input/Output Buffers: Complete buffering is provided for both input and output functions. All such operations can be completed by the computer in one interlace time (0.91 millisecond).

Parity Check: On all data transfers from main memory. Operator can elect to have parity error stop computation, or, in the case of instructions, to cause automatic reread.

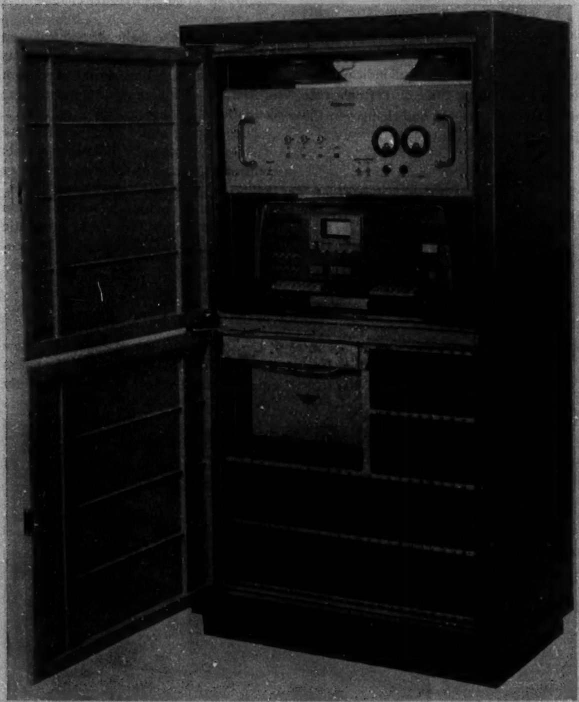


Fig. 2. LN 3000 digital computer

Protected Memory: Any multiple of 4,000 words can be protected in main memory by disabling the associated "write" amplifier. Programs stored in this section cannot be altered by random errors.

As has been mentioned previously in this paper, system reliability is obtained not only through reliable circuits, but through reliable application techniques. We have seen how the system has been designed to accommodate removal of the computer from on-line operation. Analogously, there are programming techniques available which will enhance the reliability of the computer itself. Many of these techniques have been previously reported by Ralston[1] and include such items as reasonableness and redundancy checks, detection of incorrect program jumps and endless looping, and arithmetic checks. Protected programs, written in accordance with such principles, enable an on-line control computer to compensate for most types of transient errors.

Computer Output

There are two basic digital outputs from the computer. One output is used to control and transfer data to the printers. The second output is used to control and transfer data for all control functions: addressing scanners, actuating annunciator lights, actuating on-off outputs, and adjusting analog outputs.

Printers

Each system must have a computer printer which is a standard Flexowriter with a paper tape reader and punch. This printer will be used in the preparation of computer program tapes. Programs will be entered into the computer via the paper tape reader.

The number of logging printers will vary according to system requirements. In the majority of applications, two printers will be sufficient; one printer for scheduled and on-demand printouts, and one printer for off-normal logging. Experience has indicated that the majority of maintenance problems with printers are associated with timing and adjustment of feedback contacts. In order to avoid this problem, transistorized electronic timing circuits are used for print control.

Annunciator Lights

In addition to the annunciator lights actuated by external alarm detection, 255 lights can be actuated directly from the computer. These lights will indicate such items as operating guide off-normals, rate of change off-normals, and changes in the mode of process control.

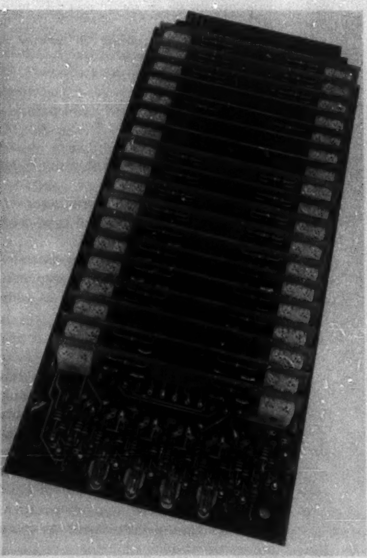


Fig. 3 (left). Logic card

Analog Outputs—Control Set Point Readjustment

Analog outputs are available to provide process control by the readjustment of local analog control loops. To obtain the maximum reliability a stepping motor is used to adjust each local analog controller. The stepping motor will normally drive a potentiometer which serves as a controller reference. The stepping motor was chosen as the reference device because of the memory inherently built in the motor. In case of power or system failures the stepping motor will remain fixed. A maximum of 255 stepping motors can be supplied with a standard system. Each motor can be addressed directly by the computer.

When a motor is selected, a pulse train is developed to rotate the motor and increase or decrease the control set point as required. The circuitry for controlling stepping motors has been designed to enforce a "reasonableness" criterion upon the change demanded by the computer. This is accomplished by limiting the magnitude of an individual change to a 1% increment.

As a further check the set point of each local controller is fed into the system as an analog input. This permits a recheck of the set point readjustment by the computer.

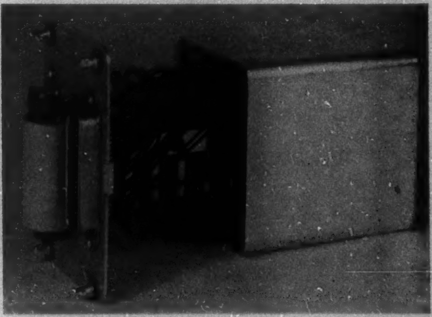
The stepping motor operates at a rate of 60 steps per second. Buffering permits continuous computer operation while set points are being adjusted.

Provision is made to allow the process operator to manually override automatic set point control. The computer is automatically informed of the manual override.

On-Off Outputs

On-off outputs are available to allow the computer to start and stop process equipment. The system will normally

Fig. 4 (below). Low-level input scanning relay



handle 255 on-off outputs, each of which can be addressed by the computer. Nonvolatile memory and manual override are also provided with each on-off output.

The circuitry for actuating on-off outputs has been designed to be fail-safe. Two computer signals are required in sequence for an on-off to be actuated. One signal is a set, the second signal is an actuate signal. The same logical circuitry is used for both the set and actuate signal. The state of logic circuitry for the set signal is inverted from the state of circuitry during the actuate signal. Unless all the logic circuitry can assume both the on and off states, no on-off output will be actuated.

System Packaging

The system is housed in cabinets 58 inches high, 37 inches wide, and 19 inches deep. The computer occupies one cabinet as shown in Fig. 2. The input-output equipment will be housed in two to five identical cabinets according to system requirements.

A maintenance and programming control panel, visible in the upper section of the computer in Fig. 2, is locked inside the computer during normal operation. The control panel may be placed on the desk next to the computer during programming and maintenance. The system control panel is similar in construction to the programming control panel and may be located at the customer's convenience.

The weight of the computer is approximately 400 pounds. Each input-output console will weigh about 250 pounds.

The equipment has been designed to operate over an ambient temperature range of 59–104 degrees Fahrenheit. Air conditioning will be incorporated for higher ambient temperatures. All components of the system are designed to be of plug-in design. A typical logic card is shown in Fig. 3. A low level input signal scanning relay is shown in Fig. 4.

Conclusions

The importance of over-all system design in the successful application of computer control systems has been emphasized. In designing such a system an appropriate balance must be achieved between the conflicting parameters of capability, reliability, and cost.

Capability specifications, such as computing speeds and the like, are relatively simply stated and compared. Reliability is more difficult to specify and compare, and in the last analysis is fully determined only after installation. For this reason restraint must be exercised by the designer in not providing an excess of capability which shows, at the expense of reliability which counts.

Attainment of satisfactory reliability has been shown to be dependent upon proper circuit design in the measurement, logic and control areas, programming techniques especially developed for online control applications, and finally a multimode system design which provides less comprehensive but adequate operation in the event that failure should occur in some of the more complex parts of the system.

References

1. **Error Detection and Error Correction in Real-Time Digital Computers, Anthony Ralston.** *Proceedings*, Western Joint Computer Conference, Institute of Radio Engineers, New York, N. Y., no. 11, 1957, pp. 179–88.

2. **The Outlook for Computer Control, I. M. Stein.** *Electrical Engineering*, vol. 78, no. 10, Oct. 1959, pp. 990–95.

Analysis and Optimization of the Bridge Magnetic Servoamplifier with True Motor Load

J. M. WALSH
ASSOCIATE MEMBER AIEE

Z. H. MEIKSIN
ASSOCIATE MEMBER AIEE

Synopsis: This paper examines the operation of a magnetic servoamplifier in which the load circuit includes the resistance and inductance of a servomotor control winding shunted by a tuning capacitor. Amplifier parameters (tuning capacitor C, limiting resistor R_s, and quiescent firing angle θ) are being optimized for best linearity and maximum motor torque. Amplifier motor transfer characteristics are established. The results are illustrated by a numerical example and verified by experiments.

A NUMBER of technical papers[1–3] have been published on the subject of the widely used Lufcy half-wave bridge magnetic servoamplifier; see Fig. 1. In none of these was a theoretical analysis performed which considered explicitly the effect of a reactive load on the operation of the amplifier. Although these papers provide valuable practical information within the limits of the approximations made, they fail to reveal characteristics of the amplifier which depend directly on the nonresistive character of the load.

Under the assumption of resistive load it may be shown that the maximum half-wave average output voltage of the amplifier cannot exceed the half-wave average supply voltage, independent of the value of R_s. This limitation results from the restricted limiter reset volt-seconds available during the reset half-cycle, equal to the supply voltage volt-seconds. A major application of the Lufcy amplifier is in conjunction with a 2-phase servomotor load. It is known from general practice that with a 2-phase servomotor load, the half-wave average output voltage of the amplifier may be made to exceed the half-wave average supply voltage, if proper adjustment of the limiting resistor is made. This is desirable in order to increase the efficiency of the amplifier, and to increase the maximum torque available from the motor amplifier set.

It is common to design and build the amplifier initially with a value of R_s approximately equal to the tuned impedance of the control phase of the servomotor, and then to decrease R_s on the breadboard gradually until no further increase in output torque of the motor for a given control signal into the amplifier is obtained. It should be noted that

---

Paper 60-784, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting, Atlantic City, N. J., June 19–24, 1960. Manuscript submitted November 13, 1959; made available for printing March 29, 1960.

J. M. Walsh is with the J. W. Fecker Division of the American Optical Company, Pittsburgh, Pa., and Z. H. Meiksin is with the University of Pittsburgh, Pittsburgh, Pa.

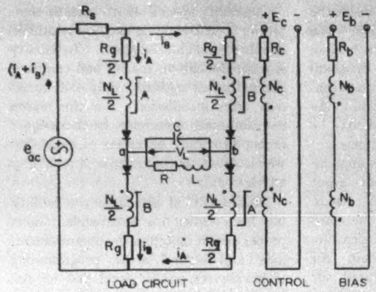


Fig. 1. Half-wave push-pull bridge magnetic servoamplifier

Table I. Optimization Equations

Parameter	Equation	Equation No. in Test
C	$C=L/[R^2+(\omega L)^2]$	26
R_s	$\frac{1}{2}\left(2\frac{V_m}{\omega}+K_{11}\right)-K_9+K_{10}$	29
θ	θ optimum $=\cos^{-1}\left[1-\frac{\omega}{V_m}(K_9+K_{10})\right]$	32

C = tuning capacitor.
R_s = limiting resistor.
θ = quiescent firing angle.

not only is this design method inefficient, but also that it does not provide an optimum design of the amplifier, because the optimum number of the gate turns, as well as other design parameters of the amplifier, depend on the value of R_s, which was unknown at the time of the design.

One of the objectives of this paper is to present an analysis which will enable one to determine the optimum value of R_s analytically and to calculate the maximum output voltage available. With these values known, an optimum amplifier design may be carried forth.

In the course of this paper expressions for amplifier voltages and motor currents as well as motor output torque will be obtained. This will lead to the analytical establishment of the transfer curve of the amplifier with the actual tuned motor load.

It will be further shown that the best linearity of the transfer curve is obtained for quiescent firing angles, θ (i.e., the firing angle of both cores when no signal is applied), greater than 90 degrees. This in turn is favorable in the light of the heat dissipation in the amplifier, which decreases with the increase of the quiescent firing angle.

Very simple solutions for waveforms and transfer curves are obtained for magnetic amplifiers in which the output current is limited by resistance only. The analysis becomes much more elaborate if this restriction is removed. Yet the final equations are practicable and useful. The equations required for the actual design of the amplifier are summarized in Table I. The application of the results are illustrated in a numerical example.

Operation

The circuit under consideration, shown in Fig. 1, consists of load, control, and bias circuits. Common to all three circuits are two magnetic cores of senstive rectangular loop material. Each core has two gate windings, a control winding, and a bias winding.

The bias circuit is designed so that with zero control signal, both cores saturate at the same angle of the a-c supply voltage, leaving the bridge balanced, since the potentials at points *a* and *b* are equal. For a control signal with the polarity shown in Fig. 1, core *A* fires before core *B*, unbalancing the bridge with the potential at point *a* more positive than at point *b*. For a polarity reversal of the control signal core *B* fires before core *A*, causing *b* to be positive with respect to point *a*.

The load circuit consists of the a-c resistance of the motor control winding *R*, the inductance of the motor control winding *L*, and the tuning capacitor *C*.

The alternating voltage source $e_{ac}=V_m \sin \omega t$, is shown with R_s. The quantity $R_g/2$ represents the sum of the gate winding resistance and rectifier forward resistance for each leg of the bridge, assuming a constant rectifier forward resistance.

Derivation of Equations

Equations for the load voltage and current, the core winding voltage, and the output torque of the motor will be derived. For the winding polarities indicated in Fig. 1, core *A* will saturate at an angle $\omega t=\alpha_1$ and core *B* will saturate at an angle $\omega t=\alpha_2$, where $\alpha_1<\alpha_2$. Accordingly, the analysis of operation is based on four distinct time intervals or periods:

Period 1: $0\leq\omega t<\alpha_1$
Period 2: $\alpha_1\leq\omega t<\alpha_2$
Period 3: $\alpha_2\leq\omega t<\pi$
Period 4: $\pi\leq\omega t<2\pi$

The intial conditions for each interval are established by the final conditions of the previous interval.

Period 1

Period 1 follows the reset half-cycle of the amplifier, period 4. During the reset half-cycle there exist voltages across and currents in the load components. No active elements being present in the load circuit during this half-cycle, these voltages and currents must eventually reduce to zero because of the dissipation in the resistance of the circuit. It will be assumed that these voltages and currents have reduced sufficiently by the end of the reset half-cycle to permit the assumption that they are equal to zero at the beginning of period 1. This assumption is in agreement with the relative magnitudes of the resistance, inductance, and capacitance in practical servomotors.

Kirchhoff's voltage equations for the circuit are then

$$V_{ac}=(i_A+i_B)R_s+N_L\frac{d\phi A}{dt}+i_AR_g \tag{1}$$

and

$$V_{ac}=(i_A+i_B)R_s+N_L\frac{d\phi B}{dt}+i_BR_g \tag{2}$$

Assuming that the circuit parameters are such that during this nonsaturation interval

$$(i_A+i_B)R_s+i_AR_g\ll N_L\frac{d\phi A}{d}$$

and

$$(i_A+i_B)R_s+i_BR_g\ll N_L\frac{d\phi B}{dt}$$

equations 1 and 2 lead to the relation

$$N_L\frac{d\phi A}{dt}=N_L\frac{d\phi B}{dt}$$

or, upon integration between the limits ϕ_{PA} and ϕ_s and between the limits ϕ_{PB} and ϕ_B respectively,

$$(\phi_s-\phi_{PA})=(\phi_B-\phi_{PB}) \tag{3}$$

The load voltage v_L is zero during this period since the potentials at points *a* and *b* of the bridge circuit are equal, assuming that the magnetizing currents of cores *A* and *B* are the same.

PERIOD 2

At the instant $t=\alpha_1/\omega$, when core *A* saturates, the load circuit of Fig. 1 becomes as shown in Fig. 2(A).

Since, during the interval under discussion, core *B* is still unsaturated, and therefore $i_B << i_{load}$, Fig. 2(A) may be simplified as shown in Fig. 2(B). The alternating voltage source is represented by

$$V_m \sin \omega t \times U\left(t-\frac{\alpha 1}{\omega}\right)$$

The unit step function

$$U\left(t-\frac{\alpha 1}{\omega}\right)$$

enters into the expression, since the voltage is switched on at the instant

$$t=\frac{\alpha 1}{\omega}$$

by the firing of core *A*. Kirchhoff's voltage equations for the equivalent circuit are

$$V_m \sin \omega t \times U\left(t-\frac{\alpha 1}{\omega}\right) = (i_L+i_C)R_{sg}+Ri_L+L\frac{di}{dt}L \quad (4)$$

and

$$V_m \sin \omega t \times U\left(t-\frac{\alpha 1}{\omega}\right) = (i_L+i_C)R_{sg}+\frac{1}{C}\int i_C dt \quad (5)$$

Since the current through an inductance and the voltage across a capacitor cannot change instantaneously, the initial conditions for this period are

$$i_L\left(\frac{\alpha 1}{\omega}\right)=0 \quad (6)$$

$$V_c\left(\frac{\alpha 1}{\omega}\right)=0 \quad (7)$$

The solution of equations 4 and 5 together with initial conditions of equations 6 and 7 gives for the motor control winding current, for $\alpha_1 \leq \omega t < \alpha_2$

$$i_L=K\left\{f_1 \cos(\alpha_1-\omega t)-f_2 \sin(\alpha_1-\omega t)+ f_3 \exp\left[r_1\left(t-\frac{\alpha_1}{\omega}\right)\right]-f_4 \exp\left[r_2\left(t-\frac{\alpha_1}{\omega}\right)\right]\right\} \quad (8)$$

(See Appendix I.)

The motor control winding voltage V_L, for the same interval is next obtained. Referring to Fig. 2(B),

$$V_L=i_L R+L\frac{di_L}{dt} \quad (9)$$

where i_L is expressed by equation 8, the solution of which is

$$V_L=K\left\{f_5 \cos(\alpha_1-\omega t)+f_6 \sin(\alpha_1-\omega t)+ f_7 \exp\left[r_1\left(t-\frac{\alpha_1}{\omega}\right)\right]- f_8 \exp\left[r_2\left(t-\frac{\alpha_1}{\omega}\right)\right]\right\} \quad (10)$$

(See Appendix I.)

The instantaneous capacitor current is obtained from the capacitor voltage which is identical to the motor winding voltage expressed by equation 10. For the capacitor

$$i_C=C\frac{dV_L}{dt}$$

Performing this differentiation gives

$$i_C=CK\left\{\omega f_5 \sin(\alpha_1-\omega t)-\omega f_6 \cos(\alpha_1-\omega t)+ r_1 f_7 \exp\left[r_1\left(t-\frac{\alpha_1}{\omega}\right)\right]- r_2 f_8 \exp\left[r_2\left(t-\frac{\alpha_1}{\omega}\right)\right]\right\} \quad (11)$$

which expresses the instantaneous capacitor current for the period $\alpha_1 \leq \omega t < \alpha_2$.

PERIOD 3

At instant $t=\alpha_2/\omega$, when core *B* saturates, the load circuit of Fig. 2(A) becomes as shown in Fig. 2(C).

As an aid to evaluating load voltage and current for this period, Thévenin's theorem will be applied. To this end the circuit of Fig. 2(C) is opened at points 1 and 2. The Thévenin equivalent voltage of the

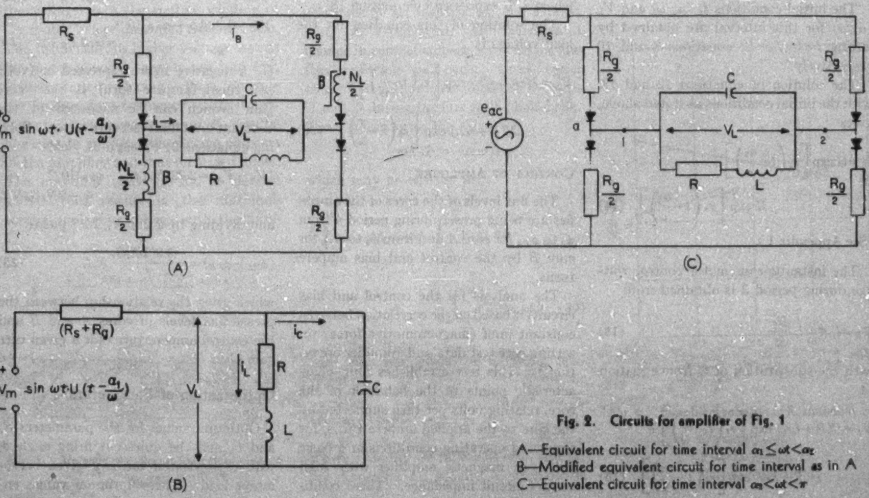


Fig. 2. Circuits for amplifier of Fig. 1

A—Equivalent circuit for time interval $\alpha_1 \leq \omega t < \alpha_2$
B—Modified equivalent circuit for time interval as in A
C—Equivalent circuit for time interval $\alpha_2 < \omega t < \pi$

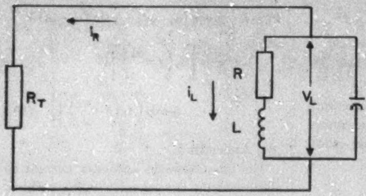


Fig. 3 (left). Thévenin's equivalent circuit for amplifier of Fig. 1 for time interval $\alpha_2 \leq \omega t < \pi$

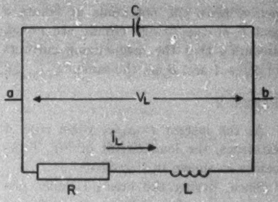


Fig. 4 (right). Load circuit for amplifier of Fig. 1 for time interval $\pi \leq \omega t < 2\pi$

circuit is equal to zero since the bridge is balanced when the load is removed. The circuit voltages during period 3 are such that all the diodes are conducting. The diodes can therefore be replaced by their forward resistances. The Thévenin equivalent resistance is thus found from Fig. 2(C) by short-circuiting the power supply e_{ac}, assuming negligible internal impedance of the power supply, by replacing the diodes by their forward resistances, and by inserting a fictitious voltage generator between points 1 and 2 after removing the load. The equivalent resistance is then found to be equal to $Rg/2$.

Thévenin's equivalent circuit of Fig. 2(C) is therefore as shown in Fig. 3.

Kirchhoff's voltage equations for the network shown in Fig. 3 are

$$\frac{1}{C}\int (i_L+i_R)dt+R_T i_R=0 \tag{12}$$

and

$$\frac{1}{C}\int (i_L+i_R)dt+Ri_L+L\frac{di_L}{dt}=0 \tag{13}$$

The initial conditions i_L (α_2/ω) and V_L (α_2/ω) for this interval are obtained by letting $t=(\alpha_2/\omega)$ in equations 8 and 10 respectively.

The solution of equations 12 and 13, with the initial conditions as stated above, gives

$$i_L=G\exp\left[r_3\left(t-\frac{\alpha_2}{\omega}\right)\right]+H\exp\left[r_4\left(t-\frac{\alpha_2}{\omega}\right)\right] \tag{14}$$

(See Appendix I.)

The instantaneous motor control voltage during period 3 is obtained from

$$V_L=i_LR+L\frac{di_L}{dt} \tag{15}$$

with the substitution of i_L from equation 14

$$V_L=G(R+Lr_3)\exp\left[r_3\left(t-\frac{\alpha_2}{\omega}\right)\right]+H(R+Lr_4)\exp\left[r_4\left(t-\frac{\alpha_2}{\omega}\right)\right] \tag{16}$$

PERIOD 4

At the instant $t=\pi/\omega$, the voltages in the circuit are such that all four diodes are blocking. The load circuit is then isolated from the rest of the amplifier components; see Fig. 4.

Kirchhoff's voltage equation for this series *R-L-C* circuit is

$$L\frac{di_L}{dt}+Ri_L+\frac{1}{C}\int i_L dt=0 \tag{17}$$

The solution of this equation with initial conditions coincident with the final conditions of period 3 is

$$i_L=M\exp\left[r_5\left(t-\frac{\pi}{\omega}\right)\right]+N\exp\left[r_6\left(t-\frac{\pi}{\omega}\right)\right] \tag{18}$$

(See Appendix I.)

Referring to Fig. 2(C), the load voltage V_L is next obtained.

$$V_L=i_LR+L\frac{di_L}{dt} \tag{19}$$

where i_L is expressed by equation 18.

The solution of this equation for the load voltage is

$$V_L=M(R+r_5L)\exp\left[r_5\left(t-\frac{\pi}{\omega}\right)\right]+N(R+r_6L)\exp\left[r_6\left(t-\frac{\pi}{\omega}\right)\right] \tag{20}$$

CONTROL OF AMPLIFIER

The flux levels of the cores of the amplifier are being preset during period 4 from ϕ_s to ϕ_{PA} for core *A* and from ϕ_s to ϕ_{PB} for core *B* by the control and bias ampere turns.

The analysis for the control and bias circuits is based on the correlation between constant mmf (magnetomotive force) resetting core test data and amplifier operation.[4] This test establishes four characteristic points of the behavior of the core, relating volts per turn supported by the core to the applied ampere turns, for simulated operating conditions of a basic half-cycle magnetic amplifier with high control circuit impedance. These conditions will be assumed to hold for the present circuit.

The voltage across the load during the gating half-cycle of the amplifier $(0<\omega t\leq\pi)$, may assume such polarity as to tend to block the rectifiers of the amplifiers. The relative magnitudes of this voltage and the supply voltage, however, are such that the rectifiers will be conducting practically until the end of the positive half-cycle of the power supply. The only period during which the amplifier is controlled is then period 4.

The totals of ampere turns applied to cores *A* and *B* during the reset half-cycle are respectively

$$AT_A=-AT_{\text{bias}}+N_cI_c \tag{21}$$

and

$$AT_B=-AT_{\text{bias}}-N_cI_c \tag{22}$$

For the linear region of the core test characteristic curve the voltages per turn supported by cores *A* and *B* are respectively

$$Vi_A=2.22f(\phi_s-\phi_{PA})10^{-8}=G_1AT_A+C \tag{23}$$

$$Vi_B=2.22f(\phi_s-\phi_{PB})10^{-8}=G_1AT_B+C \tag{24}$$

G_1, a negative figure expressed in (volts per turn)/(ampere turn) is the "core gain" which can be measured by the constant mmf flux reset core tester. Solving equations 21 through 24 yields

$$(V_B-V_A)=2.22fN_L(\phi_{PA}-\phi_{PB})10^{-8}=-2N_LG_1N_cI_c$$

and dividing by $2.22f\ N_L\ 10^{-8}$ yields

$$(\phi_{PA}-\phi_{PB})=\frac{-0.9(10^8)G_1}{f}N_cI_c \tag{25}$$

which gives the relationship between the preset flux levels of cores *A* and *B* and the control ampere turns for a given core gain G_1.

Optimization of Parameters

Optimum values for the parameters *R*, and *C*, and the quiescent firing angle θ, will be evaluated for any given servomotor load. These optimum values enable the servomotor to develop maximum torque with minimum power losses in the amplifier and also give the most linear torque-control characteristics for the amplifier-motor combination.

OPTIMIZATION OF C

The optimum value of C is that which makes the R-L-C load a maximum impedance at the fundamental frequency. This results in a high fundamental component for the motor control voltage. The optimum value of C is thus determined by the requirement for resonance of the load circuit at the fundamental frequency.

$$C=\frac{L}{R^2+(\omega L)^2} \tag{26}$$

OPTIMIZATION OF R_s

The optimum value of R_s is defined as the minimum value of R_s which does not result in the firing of core B (for control polarities as shown in Fig. 1) with maximum control signal applied to the cores as explained below. After core A fires, the supply voltage Vac is divided between R_s and the load. It is therefore clear that if R_s is increased above the certain optimum value, the voltage applied to the control winding of the motor is decreased. On the other hand, it is recognized that core B cannot support more volt-seconds during the gating half-cycle, than the amount to which it was reset during the previous half-cycle. Hence if R_s is reduced below its optimum value, core B will fire before the end of the gating half-cycle, which results in the removal of the voltage from the control winding of the motor before the end of the half-cycle. The instantaneous higher voltages across the motor winding thus do not lead to a higher torque-producing voltage, which is the fundamental frequency voltage component. A reduced value of R_s results, however, in increased circulating currents in the amplifier and thus increased losses. The optimum value of R_s, as defined above, thus results in the maximum torque-producing voltage with the minimum losses in the amplifier for a given motor.

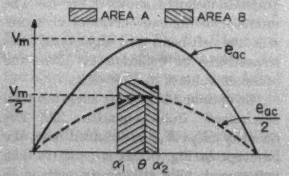


Fig. 5. Voltage waveforms for calculation of α_1, α_2, and θ

The optimum value of R_s is therefore obtained by equating the volt-seconds' change of core B during the reset half-cycle to the volt-seconds absorbed by core B during the gating half-cycle for $\alpha_1=0$ and $\alpha_2=\pi$, and solving for R_s.

Referring to Fig. 2(A),

volt-second gating_B

$$=\int_0^{\pi}\left[V_L+(i_L+i_C)\frac{Rg}{2}\right]dt \tag{27}$$

The reset volt-seconds are provided by the sum of the supply voltage and the load voltage during the reset half-cycle, as stated in equation 28:

volt-second reset_B

$$=\frac{1}{2}\int_{\frac{\pi}{\omega}}^{\frac{2\pi}{\omega}}(V_m\sin\omega t+V_L)dt \tag{28}$$

Performing the integrations indicated in equations 27 and 28 and equating the results gives

$$\frac{1}{2}\left(2\frac{V_m}{\omega}+K_{11}\right)=K_9+K_{10} \tag{29}$$

Equation 29 has R_s as the only unknown, which may be obtained graphically.

OPTIMIZATION OF θ

The maximum linearity of the amplifier is obtained for a quiescent firing angle θ, which results if $\alpha_1=0$ and $\alpha_2=\pi$ for the same control signal.[3] This value of θ is defined as the optimum value.

To determine this optimum value, a relationship between α_1, α_2, and θ must first be established. An incremental change in control ampere turns during the flux resetting half-cycle causes equal changes in volt-seconds absorbed by cores A and B during the flux gating half-cycle. Thus, referring to Fig. 5,

$$\text{area } A = \text{area } B$$

which may be written as

$$\frac{1}{2}\int_{\frac{\alpha_1}{\omega}}^{\frac{\theta}{\omega}}V_m\sin\omega t\,dt=\int_{\frac{\alpha_1}{\omega}}^{\frac{\alpha_2}{\omega}}V_B\,dt-\frac{1}{2}\int_{\frac{\alpha_1}{\omega}}^{\frac{\theta}{\omega}}V_m\sin\omega t\,dt \tag{30}$$

This equation yields the following relationship:

$$(\cos\alpha_1-\cos\theta)\frac{V_m}{\omega}=g_1+g_2 \tag{31}$$

where g_1 and g_2 are functions of α_1 and α_2; see Appendix I. Finally, θ optimum is found by letting $\alpha_1=0$ and $\alpha_2=\pi$ in equation 31

$$\theta\text{ optimum}=\cos^{-1}\left[1-\frac{\omega}{V_m}(K_9+K_{10})\right] \tag{32}$$

(See Appendix I.)

Amplifier-Motor Characteristics

The purpose of this section is to obtain a relationship between the motor torque and the amplifier control signal so that one may predict the amplifier-motor transfer characteristic curve.

The torque developed by a 2-phase servomotor, having a fixed reference voltage, is proportional to the rms value of the fundamental voltage component (of the reference voltage frequency) applied to the control phase.

$$\text{torque}=K_TV_L\text{ (fundamental)} \tag{33}$$

where K_T is the motor torque constant. The motor control voltage as a function of time V_L has been derived above for $0\leq\omega t<2\pi$, in terms of four distinct time intervals referred to as periods 1, 2, 3, and 4 respectively. This voltage during the first interval is zero, and for the remaining three intervals it is given by equations 10, 16, and 20. Let these voltages be denoted as V_{L_1}, V_{L_2}, V_{L_3}, and V_{L_4} respectively.

The motor control voltage can be expressed in terms of a Fourier series:

$$V_L=a_0+\sum_{K=1}^{K=\infty}a_k\cos K\omega t+\sum_{K=1}^{K=\infty}b_k\sin K\omega$$

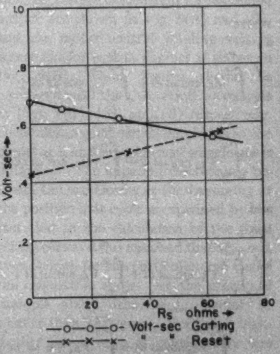
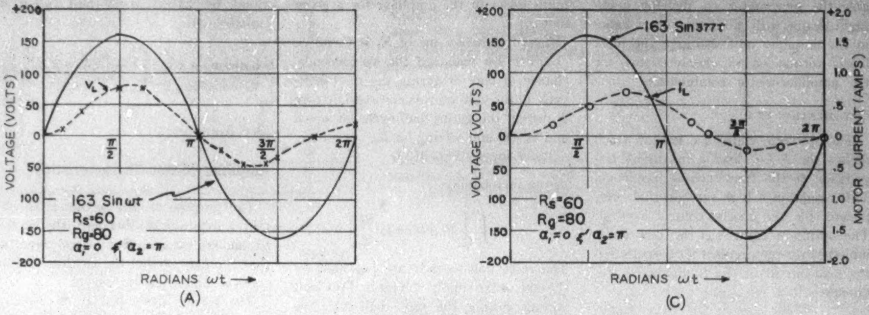


Fig. 6. Graphical solution of R_s optimum

$V_m=163$ volts
$K=377$ radians
$R_g=80$ ohms



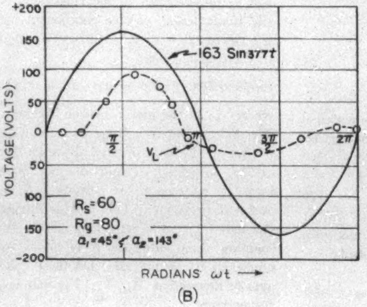


Fig. 7. Calculated load voltage waveforms

A, B—Voltage
C—Current

where the fundamental component of V_L is V_L (fundamental) $= a_1 \cos \omega t + b_1 \sin \omega t$ or

$$V_L \text{ (fundamental)} = (a_1^2+b_1^2)^{1/2} \sin\left(\omega t + \tan^{-1}\frac{a_1}{b_1}\right) \quad (34)$$

where

$a_1=$

$$\frac{\omega}{\pi}\left(\int_0^{\frac{\alpha_1}{\omega}} V_{L_1}\cos\omega t\,dt + \int_{\frac{\alpha_1}{\omega}}^{\frac{\alpha_2}{\omega}} V_{L_2}\cos\omega t\,dt + \int_{\frac{\alpha_2}{\omega}}^{\frac{\pi}{\omega}} V_{L_3}\cos\omega t\,dt + \int_{\frac{\pi}{\omega}}^{\frac{2\pi}{\omega}} V_{L_4}\cos\omega t\,dt\right)$$

and

$b_1=$

$$\frac{\omega}{\pi}\left(\int_0^{\frac{\alpha_1}{\omega}} V_{L_1}\sin\omega t\,dt + \int_{\frac{\alpha_1}{\omega}}^{\frac{\alpha_2}{\omega}} V_{L_2}\sin\omega t\,dt + \int_{\frac{\alpha_2}{\omega}}^{\frac{\pi}{\omega}} V_{L_3}\sin\omega t\,dt + \int_{\frac{\pi}{\omega}}^{\frac{2\pi}{\omega}} V_{L_4}\sin\omega t\,dt\right)$$

The evaluation of these integrals result in numerical values for a_1 and b_1. The torque of the motor is thus given by

$$\text{torque} = 0.707K_T(a_1^2+b_1^2)^{1/2} \quad (35)$$

Both a_1 and b_1 depend on the firing range α_1. To obtain the amplifier-motor transfer curve it thus remains to correlate α_1 and the control signal I_c.

Reference to equation 25 gives the correlation between the control signal I_c and the preset flux levels of the cores ϕ_{PA} and ϕ_{PB}. The correlation between the preset flux level and the firing angle α_1 is obtained as follows:

Figs. 1 and 2(A) show that for core B

$$\int_0^{\frac{\alpha_2}{\omega}} V_B dt = \int_0^{\frac{\alpha_1}{\omega}} \frac{V_{ac}}{2}dt + \int_{\frac{\alpha_1}{\omega}}^{\frac{\alpha_2}{\omega}} \left[V_L+(i_L+i_c)\frac{R_g}{2}\right]dt \quad (36)$$

but

$$V_B = \frac{N_L}{2}\frac{d\phi B}{dt}$$

whence it follows that

$$(\phi_s-\phi_{PB}) = \frac{1}{N_L}\left[\frac{V_m}{\omega}(1-\cos\alpha_1)+2(g_1+g_2)\right] \quad (37)$$

Similarly, for core A

$$\int_0^{\frac{\alpha_1}{\omega}} V_A dt = \frac{1}{2}\int_0^{\frac{\alpha_1}{\omega}} V_m \sin\omega t\,dt$$

$$V_a = \frac{NL}{2}\frac{d\phi A}{dt}$$

whence

$$(\phi_s-\phi_{PA}) = \frac{V_m}{\omega N_L}(1-\cos\alpha_1) \quad (38)$$

Subtracting equation 38 from equation 37 gives

$$\phi_{PA}-\phi_{PB} = \frac{2(g_1+g_2)}{N_L} \quad (39)$$

Combining equations 25, 31, and 39 gives

$$\alpha_1 = \cos^{-1}\left(-\frac{0.9G_1N_LN_cI_c}{V_M}+\cos\theta\right) \quad (40)$$

which yields the necessary relationship between α_1 and I_c.

Numerical Example

For the motor listed in Appendix II the measured a-c resistance of the motor control winding R was 96.5 ohms and its inductance L was 0.245 henry. The motor tuning capacitor is calculated from equation 26 and is equal to 13.7 microfarads.

The optimum value of R_s is obtained from the transcendental equation 29 by graphical means, as shown in Fig. 6. The intercept of the two curves which represent the left-hand and right-hand sides respectively of equation 29 gives the optimum value of R_s as 60 ohms.

The optimum bias firing angle θ is then determined from equation 32. The quantity (K_9+K_{10}) is obtained from the intercept in Fig. 6 and is equal to 0.55 volt-second. For the supply voltage as detailed in Appendix II, $V_m=163$ volts and $\omega=377$ radians per second. The op-

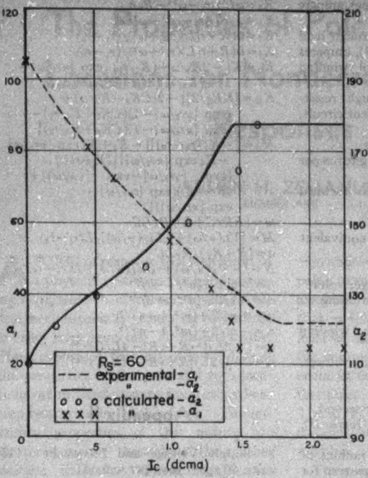


Fig. 8 (left). Calculated and experimental core firing angles versus control current

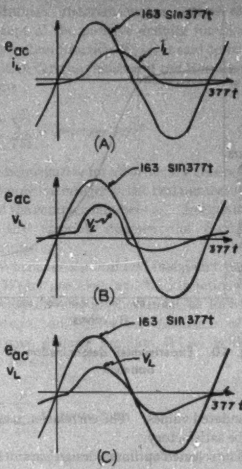


Fig. 9 (right). Experimental waveforms. (Direct tracings from oscilloscope photographs)

A—Load current at full output

B—Load voltage for $\alpha_1=45$ degrees

C—Load voltage at full output

timum value of θ is thus determined as 105 degrees.

The number of gate turns required for a selected core is then determined to permit supporting the amount of volt-seconds as obtained from equation 36 after letting $\alpha_1=0$ and $\alpha_2=\pi$. This yields 1,620 load turns.

The control and bias circuits are designed by the conventional approach based on the core test data information.

No further waveforms need be calculated for the design of the amplifier. A number will be, however, in order to compare the theoretically derived equations with experimental results.

The motor control voltage waveform at maximum output which is calculated from equations 10 and 20 is shown in Fig. 7(A). That which is calculated from equations 8 and 18 is shown in Fig. 7(C).

The motor control voltage waveform for $\alpha_1=45$ degrees is calculated from equations 10, 16, and 20, and shown in Fig. 7(B). The appropriate value of $\alpha_2=143$ degrees was obtained from equation 31.

The firing angle α_1 was calculated for various values of the control current I_c from equation 40; see Fig. 8. The corresponding values of the firing angle α_2, for the above values of α_1, were then found from equation 31 and are also shown in Fig. 8.

A word of caution is in order with regard to the application of equation 40. This equation implies that a quiescent firing angle θ is obtained with zero control ampere turns, as established by the bias ampere turns N_BI_B. The equation further implies that a value for N_cI_c can be found for which $\alpha_1=0$. This will happen when the bias ampere turns are completely counterbalanced by the control ampere turns. The net reset ampere turns for the core are then zero. Such a value of $\alpha_1=0$ would be obtained for a core with a perfect square hysteresis loop. A practical core, not having such a perfect loop, requires some volt-seconds during the gating half-cycle before saturation occurs, when zero reset ampere turns are applied to the core. Hence α_1 can never reduce to zero degrees. A minimum value for α_1 can be calculated if values of maximum flux density B_m and residual flux density B_r for the core are known. The constant mmf core tester[4] gives the necessary information of the core under the appropriate control conditions in terms of a voltage $Vi0$, which is the voltage per turn supported by the core when zero reset ampere turns are applied to it. For the particular core of the present numerical example the value was $Vi0=1.2\times10^{-3}$ volts per turn.

The minimum firing angle is obtained from the following relationship for the average voltage:

$$V_{A0}=\frac{1}{2\pi}\int_0^{\alpha_1}\frac{Vm}{2}\sin\omega t\,d\omega t$$

$$=0.45\frac{Ng}{2}Vio \quad (41)$$

For the numerical values given in Appendix III this yields the value $\alpha_{1\min}$(=21 degrees.

Experimental Investigation

A half-wave push-pull bridge magnetic amplifier, as shown in Fig. 1, was designed and constructed to allow:

1. Investigation of motor current and voltage waveforms.
2. Measurement of the amplifier-motor transfer characteristics.
3. Examination of the selected optimum design parameters.

The details of experimental hardware and data are described in Appendix II.

Actual waveforms of the motor control winding voltage and current for maximum output are shown in Fig. 9(A) and (C), and the motor control winding voltage waveform for $\alpha_1=45$ degrees is shown in Fig. 9(B). These figures should be compared with the theoretical waveforms in Figs. 7(C), (A), and (B) respectively. Generally, a good correlation between the theoretical and experimental waveforms is obtained. The apparent discrepancy between the waveforms at the beginning of the positive half-cycle is explained by the fact that in the calculation of the theoretical curves it was assumed that $\alpha_{1\min}=0$, even though for the particular core tested the calculation shows (see the numerical example above) that $\alpha_{1\min}$ cannot be smaller than 21 degrees. For many high-quality cores $\alpha_{1\min}$ is considerably less than 21. Setting $\alpha_{1\min}=0$ simplifies the calculations considerably.

Core "firing" angles α_1 and α_2 were measured for various control currents and are shown in Fig. 8, together with the

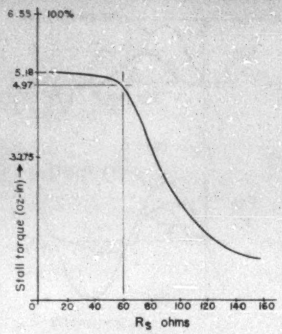


Fig. 10. Experimental determination of R_s optimum

calculated values. The correlation seems to be satisfactory.

The selected optimum design parameter R_s was examined experimentally by observing the indication of the maximum motor torque, while varying the value of R_s, as shown in Fig. 10. The maximum motor torque was observed to rise sharply as the value of R_s was decreased from a value of 150 ohms to the calculated optimum value of 60 ohms, giving 76% rated motor torque output for the latter. For further decrease in R_s, the maximum motor torque increased only very slightly, giving 79% rated motor torque output for values of R_s less than 20 ohms.

Motor output torque versus amplifier input characteristics for the various values of θ were obtained. The calculated optimum angle $\theta=105$ degrees gave the widest linear range.

Summary

The half-wave push-pull bridge magnetic amplifier with a tuned motor load has been investigated. The special behavior of the amplifier under nonresistive load lead to the optimization of amplifier design parameters. The optimization equations are summarized in Table I.

Nomenclature

$AT_A(AT_B)$ = total reset mmf for core A (core B), ampere turns
AT_{bias} = bias mmf, ampere turns
B_m = maximum or saturation magnetic flux density, gausses
B_r = residual magnetic flux
C = capacitance, farads
e_{ac} = instantaneous supply voltage
G_i = "core gain," volts per turn per ampere turn
$i_A(i_B)$ = instantaneous current through load winding of core A (core B), amperes
i_L = instantaneous motor control winding current, amperes
i_R = instantaneous current through resistance of Thévinin's equivalent circuit, amperes
N_cI_c = control mmf, ampere turns
N_L = number of total load winding turns per core
R_g = load winding resistance plus forward resistance of diode, ohms
R_s = limiting resistor, ohms
R_T = resistance of Thévinin's equivalent circuit, ohms
t = time, seconds
Vac = instantaneous supply voltage, volts
$ViA(V_{iB})$ = voltage per turn of winding on core A (core B), rms volts
$V_L(t)$ = instantaneous motor control winding voltage, volts
V_L = average motor control winding voltage, volts
V_m = peak supply voltage, volts
ω = angular frequency, radians per second
$\alpha_1(\alpha_2)$ = firing angle of core A (core B), degrees
γ = a value of $\omega t>\pi$, degrees or radians at which load voltage becomes zero for $\alpha_1=0$ and $\alpha_2=\pi$
θ = quiescent firing angle, degrees or radians
$\phi_A(\phi_B)$ = flux level of core A (core B), webers
$\phi_{PA}(\phi_{PB})$ = preset flux level of core A (core B), webers
ϕ_S = saturation flux, webers

Appendix I

Identification of constants used in the text:

$a=[L+RC(R_s+R_g)]/[LC(R_s+R_g)]$
$b=[R+(R_s+R_g)]/[LC(R_s+R_g)]$
$d=Vm/(R_s+R_g)LC$
$f_1=(r_1r_2-\omega^2)\sin\alpha_1-a\omega\cos\alpha_1$
$f_2=(r_1r_2-\omega^2)\cos\alpha_1+a\omega\sin\alpha_1$
$f_3=[(r_2^2+\omega^2)/(r_1-r_2)](\omega\cos\alpha_1+r_1\sin\alpha_1)$
$f_4=[(r_1^2+\omega^2)/(r_1-r_2)](\omega\cos\alpha_1+r_2\sin\alpha_1)$
$f_5=[R(r_1r_2-\omega^2)+\omega^2La]\sin\alpha_1+[\omega L(r_1r_2-\omega^2)-R\omega a]\cos\alpha_1$
$f_6=[\omega L(r_1r_2-\omega^2)-R\omega a]\sin\alpha_1-[R(r_1r_2-\omega^2)+\omega^2La]\cos\alpha_1$
$f_7=[(r_2^2+\omega^2)/(r_1-r_2)](R+r_1L)(\omega\cos\alpha_1+r_1\sin\alpha_1)$
$f_8=[(r_1^2+\omega^2)/(r_1-r_2)](R+r_2L)(\omega\cos\alpha_1+r_2\sin\alpha_1)$
$F_1=\cos(\alpha_1-\alpha_2)-1$
$F_2=\sin(\alpha_2-\alpha_1)$
$F_3=\exp[(\alpha_2-\alpha_1)r_1/\omega]-1$
$F_4=\exp[(\alpha_2-\alpha_1)r_2/\omega]-1$
$g_1=K(F_1f_6/\omega+F_2f_5/\omega+F_3f_7/r_1-F_4f_8/r_2)$
$g_2=KRg/2[F_1(Cf_5-f_2/\omega)+F_2(-Cf_6+f_1/\omega)+F_3(Cf_7+f_3/r_1)-F_4(Cf_8+f_4/r_2)\}$
$G=\{r_3i_L(\alpha_2/\omega)+[Li_L(\alpha_2/\omega)+R_TCV_L(\alpha_2/\omega)]/R_TLC\}/(r_3-r_4)$
$H=\{r_4i_L(\alpha_2/\omega)+[Li_L(\alpha_2/\omega)+R_TCV_L(\alpha_2/\omega)]/R_TLC\}/(r_4-r_3)$
$K=d/(r_1^2+\omega^2)(r_2^2+\omega^2)$
$K_2=r_1r_2-\omega^2$
$K_3=\omega(r_2^2+\omega^2)/(r_1-r_2)$
$K_4=\omega(r_1^2+\omega^2)/(r_1-r_2)$
$K_5=\omega L(r_1r_2-\omega^2)-R\omega a$
$K_6=-R(r_1r_2-\omega^2)-\omega^2La$
$K_7=\omega(R+r_1L)(r_2^2+\omega^2)/(r_1-r_2)$
$K_8=\omega(R+r_2L)(r_1^2+\omega^2)/(r_1-r_2)$
$K_9=K\{-2K_5/\omega+(K_7/r_1)(\exp[\pi r_1/\omega]-1)-(K_8/r_2)(\exp[\pi r_2/\omega]-1)\}$
$K_{10}=[KRg/2]\{-2(CK_5-K_2/\omega)+(\exp[\pi r_1/\omega]-1)(CK_7+K_3/r_1)-(\exp[\pi r_2/\omega]-1)(CK_8+K_4/r_2)\}$
$K_{11}=[-K/(r_5-r_6)]\{-K_1+K_3\exp[\pi r_1/\omega]-K_4\exp[\pi r_2/\omega]\}\{(R+r_5L)(\exp[\gamma r_5/\omega]-\exp[\gamma\pi/\omega)]+(R+r_6L)(\exp[r_6\gamma/\omega]-\exp[\pi r_6/\omega])\}$
$m=(RR_TC+L)/R_TLC$
$M=[V_L(\pi/\omega)+r_5Li_L(\pi/\omega)]/L(r_5-r_6)$
$n=(R+R_T)/R_TLC$
$N=[-V_L(\pi/\omega)+r_6Li_L(\pi/\omega)]/L(r_5-r_6)$
$r_1=-a/2+(a^2/4-b)^{1/2}$
$r_2=-a/2-(a^2/4-b)^{1/2}$
$r_3=-m/2+(m^2/4-n)^{1/2}$
$r_4=-m/2-(m^2/4-n)^{1/2}$
$r_5=(1/2L)[-R+(R^2-4L/C)^{1/2}]$
$r_6=(-1/2L)[R+(R^2-4L/C)^{1/2}]$

Appendix II

1. Supply Voltage and Frequency: 115 volts, 60 cps (cycles per second)

2. Magnetic Cores
   Size: 1.5-inch inside diameter, 2.5-inch outside diameter, 0.5-inch height
   Material: Hipernik V (50% nickel, 50% iron alloy, grain-oriented)
   Construction: Toroidal cores, wound from 2-mil thick strip
   Core Gain: $G_i=-0.0325$ volt per turn per ampere turn, measured at 60-cps excitation
   Windings:
      Load windings (total per core), 1,620 turns no. 29 AWG (American Wire Gage), 40 ohms
      Control winding, 500 turns no. 38 AWG
      Bias winding, 300 turns no. 38 AWG

3. Servomotor
   Diehl type *FPE 25*
   Voltage rating, 115 volts for reference phase, 57.5 volts for control phase
   Frequency, 60 cps
   Torque at stall, 6.55 inch-ounces
   No-load speed, 3,600 rpm

References

1. An Improved Magnetic Servo Amplifier, C. W. Lufcy, A. E. Schmid, P. W. Barnhart. *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 71, Sept. 1952, pp. 281–89.

2. Design Considerations of the Half-Wave Bridge Magnetic Amplifier, C. W. Lufcy, H. H. Woodson. *Ibid.*, vol. 73, July 1954, pp. 220–26.

3. Comparative Study of Magnetic Servo Amplifiers, Z. H. Meiksin. *"Proceedings*, Special Technical Conference on Nonlinear Magnetics and Magnetic Amplifiers," AIEE, Aug. 1958, pp. 158–79.

4. Magnetic Characteristics Pertinent to the Operation of Cores in Self-Saturating Magnetic Amplifiers, R. W. Roberts. *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 73, 1954 (Jan. 1955 section), pp. 682–90.

The Properties of Pole and Zero Locations for Nondecreasing Step Responses

ARMEN H. ZEMANIAN
MEMBER AIEE

A FORM of response to a step input in a physical system is the nondecreasing type, which is of interest to the circuit designer, especially when an overshoot in the step response is objectionable. The purpose of this paper is to discuss the properties of the pole and zero locations of rational system functions whose step responses are of this form. The few known properties of this nature are scattered mainly throughout the mathematical literature. These results are gathered here and presented in a unified manner. Secondly, several other results are developed that are believed to be new. In some cases, the proofs of the following theorems are merely sketched or omitted entirely when they are available in the referenced literature.

This paper is arranged into three parts. In the first part, two well-known theorems from the theory of Laplace transforms are stated, each of which gives a necessary and sufficient condition on the system function that insures a nondecreasing step response. These conditions apply whether or not the system function is rational. In the second part, necessary conditions that the pole and zero locations must satisfy for a nondecreasing step response are given. The third part is concerned with sufficient, but not necessary, conditions on the pole and zero locations that insure a nondecreasing step response. This part constitutes the major portion of this paper and is especially of interest in the design of systems.

Only those rational system functions, whose corresponding responses to input step functions are bounded, will be considered. In particular, the discussion will be restricted to system functions $W(s)$ having the form,

$$W(s)=K\frac{\prod_{i=1}^{h}(s+\eta_i)\prod_{i=1}^{2g}(s+\nu_i)}{\prod_{i=1}^{m}(s+\rho_i)\prod_{i=1}^{2q}(s+\xi_i)} \qquad (1)$$

where $s=\sigma+j\omega$ denotes the complex frequency variable; $-\eta_i$ and $-\nu_i$ denote the real and complex zeros, respectively; $-\rho_i$ and $-\xi_i$ denote the real and complex poles, respectively; and K is a positive constant. It will be assumed throughout that $m+2q>h+2g$ so that there are at least as many poles as zeros. Moreover, the complex poles and zeros of $W(s)$ will occur in complex conjugate pairs and all the poles will have negative real parts. The η_i, ν_i, ρ_i, and ξ_i need not be distinct so that multiple poles and zeros are possible.

The response $a(t)$ to a unit step input function is related to the system function $W(s)$ through the following Laplace-Stieltjes transform.

$$W(s)=\int_0^\infty e^{-st}\,da(t) \qquad (2)$$

As a result of the aforementioned restrictions on $W(s)$, $a(t)$ will be a sum of terms having the form

$$Ct^{\mu-1}e^{-\alpha t}\cos(\beta t+\psi) \qquad (3)$$

where C, α, β, and ψ are real constants with $\alpha>0$ and where μ is a positive integer. In addition, $a(t)$ may also have a constant term. Thus, $a(t)$ is a differentiable function for $t>0$. The condition that $a(t)$ is nondecreasing for $t>0$ is entirely equivalent to the statement that the unit impulse response $w(t)$, which is the derivative of $a(t)$, is nonnegative for $t>0$ and has possibly a positive impulse at $t=0$.

Part I. Necessary and Sufficient Conditions

In a previous paper[1] two realizability theorems were given which insured that an impulse response matrix was obtainable from a linear, fixed, passive n-terminal network. These results were based upon two theorems from Laplace transform theory, Mathias' theorem[2] and Bernstein's theorem; see also theorem 12a, page 160 of Widder.[3] Because of the similarity between the direct and inverse Fourier transforms, these same two theorems may also be applied to determine system functions corresponding to nondecreasing step responses.

Mathias' theorem makes use of the concept of the positive definite function. The system function $W(j\omega)$ is said to be a positive definite function if the Hermitian forms

$$\sum_{i=1}^{p}\sum_{k=1}^{p}W(j\omega_i-j\omega_k)y_iy_k^* \qquad (4)$$

are nonnegative for all arbitrary choices of the positive integer p, the real numbers ω_i, and the complex numbers y_i. In equation 4, the asterisk denotes the complex conjugate.

Whenever the real and imaginary parts of $W(j\omega)$ are even and odd functions of ω, respectively, as they must be for real systems, an equivalent definition of a positive definite function $W(j\omega)$ is that the forms,

$$\sum_{i=1}^{p}\sum_{k=1}^{p}R(\omega_i-\omega_k)x_ix_k,$$

are nonnegative for all arbitrary choices of the positive integer p and the real numbers ω_i and x_i where $R(\omega)$ denotes the real part of $W(j\omega)$.

Under the aforementioned restrictions on $W(j\omega)$, a generalization of Mathias' theorem, which is due to Bochner,[4] may be stated as follows:

Bochner's Theorem: A necessary and sufficient condition for $a(t)$ to be nondecreasing is that $W(j\omega)$ be a positive definite function of ω.

Another such condition is given by the theorem due to Bernstein concerning completely monotonic functions. The real function $W(\sigma)$ of the real variable σ is said to be completely monotonic for $\sigma>0$ if $W(0)=W(0+)$ and if

$$(-1)^kW^{(k)}(\sigma)\geq 0 \qquad (5)$$

for $\sigma>0$ and for all nonnegative integers k. In equation 5, $W^{(k)}(\sigma)$ denotes the kth derivative of $W(\sigma)$. With this notation Bernstein's theorem is the following:

Bernstein's Theorem: A necessary and sufficient condition for $a(t)$ to be nondecreasing is that $W(\sigma)$ be a completely monotonic function for $\sigma>0$.

That Bernstein's theorem may be applied to system functions having nondecreasing step responses has previously been pointed out by Choudhury and

---

Paper 60-782, recommended by the AIEE Basic Sciences Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting, Atlantic City, N. J., June 19–24, 1960. Manuscript submitted December 14, 1959; made available for printing March 29, 1960.

ARMEN H. ZEMANIAN is with New York University, New York, N. Y.

The author wishes to acknowledge the advice of Prof. C. F. Rehberg.

Chakrabarti[5] and by Aaron and Segers.[6] As mentioned previously, these theorems hold for nonrational system functions as well as rational ones. There are several other necessary and sufficient conditions of this type. In particular, the results of Cramér[7] and Khintchine[8] should be noted.

The rest of this paper will be concerned with the pole and zero locations of rational system functions. Necessary and sufficient conditions on these locations for a nondecreasing $a(t)$ have been derived by Macdiarmid[9] and Cutteridge[10] in the special cases where the system function has two or three poles and at most two or three zeros, respectively. The present investigation will consider sufficient conditions on the poles and zeros of $W(s)$, which insure that $a(t)$ is nondecreasing, in the general case where the number of poles and zeros of $W(s)$ is unrestricted. First, however, some necessary conditions for a nondecreasing $a(t)$ will be discussed.

Part II. Necessary Conditions on the Pole and Zero Locations

A set of necessary conditions that was first obtained by Mulligan[11] and subsequently by Lukacs and Szasz[12] and by Takano[13] is given by the following statement:

Mulligan's Theorem: The following conditions on $W(s)$ must hold if $a(t)$ is a nondecreasing function.

1. If $W(s)$ has a complex pole at $-\alpha+j\beta$ where $\alpha>0$, then $W(s)$ has at least one real pole at $-\rho$ such that $\alpha>\rho\geq 0$.
2. If the real pole closest to the imaginary axis occurs at $-\rho$ and if there are complex poles at $-\rho\pm j\beta_i$ $(i=1, 2, \ldots, N)$, then no such complex pole can have a multiplicity greater than the multiplicity of the real pole at $-\rho$.
3. If the poles closest to the imaginary axis occur only at the three points $-\rho\pm j\beta$ and $-\rho$ (so that $N=1$), and if these poles have the same multiplicity μ, then $A\geq 2|B|$ where A is the coefficient of $(s+\rho)^{-\mu}$ and B is the coefficient of $(s+\rho+j\beta)^{-\mu}$ in the partial fraction expansion of $W(s)$.

These necessary conditions may be established by expanding $W(s)$ into partial fractions and then taking the inverse Laplace transformation to get the unit impulse response $w(t)$. If the poles closest to the imaginary axis are all complex, so that condition 1 is not fulfilled, then the asymptotic behavior of $w(t)$ as t increases indefinitely is given by a sum of terms of the form 3 with no constant term. It follows that $w(t)$ must take on negative values so that $a(t)$ cannot be nondecreasing. Similarly, to establish condition 2, consider the case where there are complex poles whose real parts are equal to the real pole closest to the imaginary axis. For each pair of complex conjugate poles in this set of poles, the corresponding terms in $w(t)$ will again be of the form 3, where μ represents the multiplicity of any such pole pair. Hence, the terms in $w(t)$ that dominate as t approaches infinity will be those due to the poles of highest multiplicity in this set of poles. If the real pole is of lower multiplicity than any such complex pole pair, then the asymptotic behavior of $w(t)$ will again be given by terms of the form 3 without any constant term. This again implies that $a(t)$ cannot be nondecreasing. Finally, condition 3 is the special case of condition 2, where $N=1$. In this case the asymptotic behavior of $w(t)$ as t approaches infinity is given by

$$t^{\mu-1}e^{-\rho t}[A+2|B|\cos(\beta t+\psi)]$$

where A and B are defined in condition 3. For this expression to be nonnegative it is necessary that $A>2|B|$.

If we restrict ourselves only to those rational system functions that have no zeros in the finite s plane, then Mulligan's conditions are also sufficient for a nondecreasing step response when the number of poles is no greater than four. This will be demonstrated later. However, his conditions are not sufficient when the number of poles equals five. Takano[13] gives a counterexample.

Other necessary conditions may be derived using the approach given on page 167 of Widder.[3] Let $a(t)$ be nondecreasing and not a constant for $0<t<\infty$. Since

$$(-1)^k W^{(k)}(\sigma)=\int_0^\infty t^k e^{-\sigma t}da(t)$$

$(-1)^k W^{(k)}(\sigma)$ is positive along the entire real axis within the region of convergence for this integral. In other words, if the largest real part of all the poles of $W(s)$ is denoted by $-\delta$, then

$$(-1)^k W^{(k)}(\sigma)>0 \tag{6}$$

for $-\delta<\sigma<\infty$.

Now consider

$$-\frac{d}{d\sigma}\log W(\sigma)=-\frac{W^{(1)}(\sigma)}{W(\sigma)} \tag{7}$$

and

$$\frac{d^2}{d\sigma^2}\log W(\sigma)=\frac{W(\sigma)W^{(2)}(\sigma)-[W^{(1)}(\sigma)]^2}{[W(\sigma)]^2} \tag{8}$$

Because of equation 6, equation 7 is positive for $\sigma>-\delta$. Furthermore, for x_1 and x_2 real and for $\sigma>-\delta$,

$$x_1^2W(\sigma)-2x_1x_2W^{(1)}(\sigma)+x_2^2W^{(2)}(\sigma)=\int_0^\infty (x_1+x_2t)^2e^{-\sigma t}da(t)\geq 0$$

Equality holds only if $x_1=x_2=0$. The fact that this quadratic form is positive definitely implies that, for $\sigma>-\delta$, the numerator of the right hand side of equation 8 is positive so that equation 8 itself is positive. Substituting equation 1 into equations 7 and 8, one obtains:

Widder's Theorem: If $a(t)$ is nondecreasing, then the poles and zeros of $W(s)$ must satisfy the following conditions:

$$\sum_{i=1}^{h}\frac{1}{(\sigma+\eta_i)^k}+\sum_{i=1}^{2g}\frac{1}{(\sigma+\nu_i)^k}<\sum_{i=1}^{m}\frac{1}{(\sigma+\rho_i)^k}+\sum_{i=1}^{2q}\frac{1}{(\sigma+\xi_i)^k} \tag{9}$$

for $k=1$ and 2 and for $\sigma>-\delta$.

As an immediate consequence of this theorem, we have:

Widder's Corollary: If $a(t)$ is nondecreasing, all the real zeros of $W(s)$ must lie to the left of the largest real pole of $W(s)$.

This fact may also be developed from Bochner's theorem as follows. The Laplace-Stieltjes transform, equation 2, may be written as

$$W(\sigma+j\omega)=a(0+)+\int_0^\infty e^{-j\omega t}e^{-\sigma t}w(t)dt \tag{10}$$

where $a(0+)\geq 0$. Now, $e^{-\sigma t}w(t)$ is nonnegative and hence its Fourier transform is a positive definite function by Bochner's theorem. Moreover, the addition of a positive constant will not destroy the positive definite character. Hence, $W(\sigma+j\omega)$ is a positive definite function of ω. A property of such functions is that they are positive at $\omega=0$, except in the trivial case where the function is identically zero, as can be seen by setting $p=1$ in equation 3. Thus, $W(\sigma)>0$ for $\sigma>-\delta$ so that $W(\sigma)$ has no zeros for $\sigma>-\delta$.

Actually, Widder's corollary is a special case of a result due to Laguerre and extended by Widder,[14] which states that the number of changes of sign of $w(t)$ cannot be less than the number of real zeros of $W(s)$ within the region of convergence of the Laplace transformation. This was pointed out to the author by Professor C. F. Rehberg.

Finally, in the case where the number of poles of $W(s)$ equals the number of its zeros $(h+2g=m+2q)$, an application of the initial and final value theorems to equation 1 yields still another necessary condition for a nondecreasing step response. Namely,

$$\prod_{i=1}^{h} \eta_i \prod_{i=1}^{2g} \nu_i > \prod_{i=1}^{m} \rho_i \prod_{i=1}^{2q} \xi_i$$

Part III. Sufficient Conditions on the Pole and Zero Locations

Sufficient Conditions That Are Obtainable from Elementary Rational Functions. In this and the following sections, various sets of sufficient conditions on rational system functions that insure nondecreasing step responses will be presented. The convolution theorem, theorem 11.7b, page 91 of Widder,[3] may then be used to apply these results to a much wider class of system functions. For instance, taking the product of two system functions, $W_1(s)$ and $W_2(s)$, corresponds to taking the convolution of their two step responses, $a_1(t)$ and $a_2(t)$. That is,

$$\mathfrak{L}^{-1}[W_1(s)W_2(s)] = a_1(t)a_2(0+) + \int_0^t + a_1(t-x)da_2(x) \qquad (11)$$

Hence, if $a_1(t)$ and $a_2(t)$ are both nondecreasing, the step response corresponding to the product of $W_1(s)$ and $W_2(s)$ is also nondecreasing. The same result holds for a product of any number of such system functions.

As an example, consider the elementary rational functions given by equations 12 through 16. The step response of equation 12 is nondecreasing and those of equations 13 through 16 will also be nondecreasing if the corresponding inequalities on the pole and zero locations are satisfied.[10]

$$\frac{1}{s+\rho} \qquad (12)$$

$$\frac{s+\eta}{s+\rho}, \quad \eta > \rho \qquad (13)$$

$$\frac{1}{(s+\rho)(s+\xi)(s+\xi^*)}, \; Re\xi \geq \rho \qquad (14)$$

$$\frac{(s+\eta_1)(s+\eta_2)}{(s+\rho_1)(s+\rho_2)},$$

$$\eta_1+\eta_2 \geq \rho_1+\rho_2 \text{ and } \min(\eta_1, \eta_2) \geq \min(\rho_1, \rho_2) \qquad (15)$$

$$\frac{(s+\nu)(s+\nu^*)}{(s+\rho_1)(s+\rho_2)}, \; 2Re\nu \geq \rho_1+\rho_2 \qquad (16)$$

Hence, if a system function is a product of such elementary functions, its step response must also be nondecreasing. It is also possible to establish results of this nature by use of Bernstein's theorem rather than by the convolution integral.[6,15] It should be noted that combining equation 12 with equation 14 shows that Mulligan's conditions are sufficient as well as necessary for a nondecreasing $a(t)$ if the system function has no zeros and no more than four poles.

As was mentioned previously, Macdiarmid[9] and Cutteridge[10] have developed a fairly complete set of necessary and sufficient conditions on the pole and zero locations for rational functions having no more than three poles and three zeros. Of course, any one of these elementary functions may be used as factors for system functions as was done in the previous paragraph. The rest of this paper will consider transfer functions where the number of poles is not in general restricted and, furthermore, it will derive sufficient conditions that do not appear to be obtainable by the use of the convolution integral from conditions on elementary functions.

Sufficient Conditions for a General System Function. The next theorem[16] makes use of the following notation and numbering system for the poles and zeros. The real poles will be numbered according to their increasing magnitudes; that is, $0 < \rho_1 \leq \rho_2 \leq \ldots \leq \rho_m$. All the zeros and complex poles will be denoted by $-\gamma_i$ and numbered according to their increasing real parts; that is, denoting the real and imaginary parts of γ_i by α_i and β_i, respectively, the γ_i are numbered according to $\alpha_1 \leq \alpha_2 \leq \ldots \leq \alpha_n$ where $n = h+2g+2q$. Those γ_i that have the same real part will be numbered in any sequence subject to this restriction. If a ρ_i or a γ_i has a multiplicity of $\mu > 1$, it will be counted μ times.

Two lemmas will be needed. Let $-G(\sigma)$ be the logarithmic derivative of $W(\sigma)$,

$$G(\sigma) = -\frac{d}{d\sigma}\log W(\sigma) = -\frac{W^{(1)}(\sigma)}{W(\sigma)} \qquad (17)$$

$$G(\sigma) = \sum_{i=1}^{m}\frac{1}{\sigma+\rho_i} + \sum_{i=1}^{2q}\frac{1}{\sigma+\xi_i} - \sum_{i=1}^{h}\frac{1}{\sigma+\eta_i} - \sum_{i=1}^{2g}\frac{1}{\sigma+\nu_i} \qquad (18)$$

$$G(\sigma) = \sum_{i=1}^{m}\frac{1}{\sigma+\rho_i} + \sum_{i=1}^{n}\frac{\pm 1}{\sigma+\gamma_i} \qquad (19)$$

In the last expression, the plus sign is used if γ_i is a pole and a minus sign is used if γ_i is a zero.

Lemma 1: If $W(s)$ has no zeros on the positive real axis or at the origin and if the $G(\sigma)$ is completely monotonic in the same interval, then $W(\sigma)$ is also completely monotonic in this interval.

Under the assumed restrictions on $W(s)$, $W(\sigma)$ is nonnegative for $\sigma \geq 0$. With this fact, the work of Aaron and Segers[6] establishes lemma 1.

Lemma 2:

If $x \geq 0$, $y_i \geq 0$, and

$$x \geq \sum_{i=1}^{n} y_i,$$

then, for $k = 1, 2, 3, \ldots,$

$$(x+1)^k + n - 1 \geq \sum_{i=1}^{n}(y_i+1)^k$$

Proof: It is evident that

$$x^k \geq \left[\sum_{i=1}^{n} y_i\right]^k \geq \sum_{i=1}^{n} y_i^k$$

Therefore

$$\sum_{i=1}^{n}(y_i+1)^k = \sum_{i=1}^{n}\left[y_i^k + \binom{k}{1}y_i^{k-1} + \binom{k}{2}y_i^{k-2} + \ldots + 1\right] \geq x^k + \binom{k}{1}x^{k-1} + \ldots + \binom{k}{k-1}x + n = (x+1)^k + n - 1$$

Theorem 1: Let $W(s)$ be a system function of the form 1 and let $m \geq n = h+2g+2q$. If $\rho_n \geq \alpha_n$ and if

$$\sum_{i=1}^{n}\alpha_i \geq \rho_1 + (n-1)\rho_n \qquad (20)$$

then the unit step response $a(t)$ is nondecreasing.

Proof: From equation 20 and the fact that the $\rho_n \geq \alpha_n$, it follows that $\alpha_1 \geq \rho_1$ so that this necessary condition is fulfilled. The inequality 20 may be rewritten as

$$\rho_n - \rho_1 \geq \sum_{i=1}^{n}(\rho_n - \alpha_i).$$

The quantities $\rho_n - \alpha_i$ are nonnegative. Thus, for $\sigma \geq 0$,

$$\frac{\rho_n-\rho_1}{\sigma+\rho_1} \geq \sum_{i=1}^{n}\frac{\rho_n-\alpha_i}{\sigma+\rho_1} \geq \sum_{i=1}^{n}\frac{\rho_n-\alpha_i}{\sigma+\alpha_i}$$

Invoking lemma 2,

$$\left(\frac{\rho_n-\rho_1}{\sigma+\rho_1}+1\right)^k + n - 1 \geq \sum_{i=1}^{n}\left(\frac{\rho_n-\alpha_i}{\sigma+\alpha_i}+1\right)^k$$

Hence,

$$\frac{1}{(\sigma+\rho_1)^k} + \frac{n-1}{(\sigma+\rho_n)^k} \geq \sum_{i=1}^{n}\frac{1}{(\sigma+\alpha_i)^k}$$

Now $\rho_n \geq \rho_i$ for $i = 1, 2, \ldots, n$, so that the last inequality is replaced by

$$\sum_{i=1}^{n}\frac{1}{(\sigma+\rho_i)^k}\geq\sum_{i=1}^{n}\frac{1}{(\sigma+\alpha_i)^k} \tag{21}$$

Moreover, for a pair of complex conjugate poles or zeros,

$$\frac{2}{(\sigma+\alpha_i)^k}>\left|\frac{1}{(\sigma+\gamma_i)^k}+\frac{1}{(\sigma+\gamma_i^*)^k}\right| \tag{22}$$

Finally, differentiating equation 18 k times, one obtains

$$(-1)^k G^{(k)}(\sigma)=k!\left[\sum_{i=1}^{m}\frac{1}{(\sigma+\rho_i)^{k+1}}+\sum_{i=1}^{2q}\frac{1}{(\sigma+\xi_i)^{k+1}}-\sum_{i=1}^{h}\frac{1}{(\sigma+\eta_i)^{k+1}}-\sum_{i=1}^{2g}\frac{1}{(\sigma+\nu_i)^{k+1}}\right] \tag{23}$$

Inserting inequalities 21 and 22 into 18 and 23, it is readily seen that $(-1)^k G^{(k)}(\sigma)$ is nonnegative for $\sigma\geq 0$ and $k=0, 1, 2, \ldots$. Thus, $G(\sigma)$ is completely monotonic in the same interval. Lemma 1 completes the proof.

As an example, consider the system function

$$\frac{(s+2.4)(s+2.4-j)(s+2.4+j)}{(s+1)(s+3)^2}$$

It satisfies the hypothesis of theorem 1 so that the corresponding step response is nondecreasing.

Sufficient Conditions When All the Poles and Zeros Have the Same Real Part. Lukacs and Szasz[17,18] have developed several sets of sufficient conditions for nondecreasing step responses for the special case where all the poles and zeros have the same real part and the imaginary parts of all the poles are integral multiples of the smallest such imaginary part. Their two simplest and, hence, most useful results will be proved in this section. The proofs given here are essentially those of Lukacs and Szasz. However, they are somewhat more elementary in that no use is made of the Vandermonde determinants.

Consider a system function having only simple poles, all of whose real parts are the same. In this section the poles will be numbered by assigning the same subscript to pairs of poles that are complex conjugates of each other. The numbering will increase in accordance with the distance of the poles from the real axis. Thus, the system function may be written as

$$W(s)=\frac{K}{(s+\alpha)\prod_{i=1}^{q}[(s+\alpha)^2+\beta_i^2]} \tag{24}$$

where $\alpha>0$ and $0<\beta_1<\beta_2<\ldots<\beta_q$. The unit impulse response $w(t)$ corresponding to this $W(s)$ is for $t>0$

$$w(t)=e^{-\alpha t}\left[K_0+\sum_{i=1}^{q}K_i\cos\beta_i t\right] \tag{25}$$

where

$$K_0=\frac{K}{\prod_{k=1}^{q}\beta_k^2}$$

$$K_i=\frac{(-1)^q K}{\beta_i^2\prod_{\substack{k=1\\k\neq i}}^{q}(\beta_i^2-\beta_k^2)}$$

From equation 25 it is seen that the unit step response $a(t)$ corresponding to this $W(s)$ will be nondecreasing if and only if the quantity

$$K_0+\sum_{i=1}^{q}K_i\cos\beta_i t$$

is nonnegative. This will be the case under the following conditions.

Lukacs-Szasz Theorem 1: If $W(s)$ is given by equation 24 with $\alpha>0$ and if $\beta_i=i\beta_1$ for $i=1, 2, \ldots, q$ (that is, if the β_i are the first q consecutive, positive, whole multiples of β_1), then $a(t)$ is nondecreasing.

Proof: It is clearly sufficient to establish the result in the case where $K=\beta_1=1$. Expanding the function $(1-\cos t)^q$ into a Fourier series, one obtains

$$(1-\cos t)^q=\sum_{i=0}^{q}c_i\cos it \tag{26}$$

where

$$c_0=\frac{2^q\Gamma\left(q+\frac{1}{2}\right)}{\sqrt{\pi}(q+1)}$$

$$c_i=2(-1)^i\frac{q(q-1)\ldots(q-i+1)}{(q+1)(q+2)\ldots(q+i)}c_0$$

Furthermore, substituting $\beta_i=i$ for $i=1, 2, \ldots, q$ into equation 25, $w(t)$ becomes

$$w(t)=e^{-\alpha t}\sum_{i=0}^{q}K_i\cos it \tag{27}$$

where

$$K_0=\frac{1}{(q!)^2}$$

$$K_i=\frac{2(-1)^i}{(q-i)!(q+i)!}$$

The theorem is proved upon comparing equations 26 and 27, and noting that $w(t)$ is the obviously nonnegative function,

$$w(t)=\frac{\sqrt{\pi}}{2^q(q!)\Gamma\left(q+\frac{1}{2}\right)}e^{-\alpha t}(1-\cos t)^q$$

Lukacs-Szasz Theorem 2: If $W(s)$ is given by equation 24 with $\alpha>0$ and if $\beta_i=(2i-1)\beta_1$ for $i=1, 2, \ldots, q$ (that is, if the β_i are the first q consecutive, odd, whole, positive multiples of β_1), then $a(t)$ is nondecreasing.

Proof: Once again, we need prove the result only for $K=\beta_1=1$. The function $(\sin t)^{2q-1}$ may be expanded into:

$$(\sin t)^{2q-1}=\sum_{i=1}^{q}b_i\sin(2i-1)t \tag{28}$$

where

$$b_i=\frac{(-1)^{i-1}(2q-1)!}{2^{2q-2}(q-i)!(q+i-1)!}$$

Now let $f(t)$ equal $w(t)/\exp(-\alpha t)$. Substituting $\beta_i=2i-1$ for $i=1, 2, \ldots, q$ into equation 25 and differentiating the resulting $f(t)$, one obtains

$$f^{(1)}(t)=\sum_{i=1}^{q}\frac{(-1)^{i-1}K}{2^{2q-2}(q-i)!(q+i-1)!}\sin(2i-1)t \tag{29}$$

Comparing equations 28 and 29, it is seen that

$$f^{(1)}(t)=\frac{K}{(2q-1)!}(\sin t)^{2q-1}$$

However, by the initial value theorem, $f(0+)=0$ for $q\geq 1$. Thus,

$$w(t)=\frac{Ke^{-\alpha t}}{(2q-1)!}\int_0^t(\sin x)^{2q-1}dx$$

This is a nonnegative function.

Using a more general approach, Lukacs and Szasz have proved other results of this nature even in the case where $W(s)$ has zeros having the same real part as that of the poles. The next theorem is an example of this. Unfortunately, most of their other results are quite involved and hence not particularly useful.

Lukacs-Szasz Theorem 3: Let $W(s)$ be of the form,

$$W(s)=K\frac{(s+\alpha)^2+\delta^2}{(s+\alpha)\prod_{i=1}^{q}[(s+\alpha)^2+\beta_i^2]}$$

where $\alpha>0$ and where $\beta_i=i\beta_1$ for $i=1, 2, 3, \ldots, q$. Then $a(t)$ is nondecreasing if and only if $2\delta^2\geq q\beta_1^2$.

Before leaving this section another technique for determining a nondecreasing step response will be noted. For this approach the zeros need not have the same real part but may lie anywhere in the s plane and be of any multiplicity. However, if all the poles are simple, have the same real part, and outnumber the zeros and if the imaginary parts of all the poles are whole multiples of the smallest, positive, imaginary part of the poles, then the unit impulse response is of the form

$$w(t)=e^{-\alpha t}\, g(t)$$

where $g(t)$ is a cosine polynomial. To determine whether $a(t)$ is nondecreasing, one need only establish the nonnegativity of $g(t)$. J. F. Schultze[19] has given a variety of criteria on the coefficients of $g(t)$ which insure that $g(t)$ is nonnegative. However, in order to use these criteria one must first determine the coefficients of $g(t)$. Hence, this approach is not as useful as the criteria on the pole and zero locations if it is these locations that are known beforehand.

The Case Where All the Poles Are Real and There Is at Least One Pair of Complex Conjugate Zeros. According to Mulligan's theorem and Widder's corollary, the real zeros and complex poles must not lie to the right of all the real poles if the step response is to be nondecreasing. However, this is not the case for the complex zeros. As a simple example, consider the system function having a pair of simple zeros at $s=\pm j\beta$ $(\beta>0)$ and a triple pole at $s=-\rho$ $(\rho>0)$.

$$W(s)=\frac{s^2+\beta^2}{(s+\rho)^3} \tag{30}$$

The corresponding unit impulse response for $t\geq 0$ is

$$w(t)=e^{-\rho t}\left[\frac{\rho^2+\beta^2}{2}\, t^2-2\rho t+1\right] \tag{31}$$

This expression is nonnegative if and only if $\beta\geq\rho$. Hence, by choosing β large enough, the impulse/response can be made nonnegative.

Brule[20] has developed a precise result of this nature in the case where all the poles are real, the pole furthest to the right being simple, and where there are no zeros other than one pair of complex conjugate zeros. He shows that, if the magnitude of the imaginary part of the zeros is larger than a certain quantity, the corresponding impulse response will be nonnegative.

The possibility of generating a nondecreasing step response by choosing the imaginary part of the complex conjugate zeros large enough exists for more general system functions. For instance, given any system function having only real poles and only complex zeros, the step response can always be made nondecreasing without changing the real parts of the zeros simply by placing the zeros far enough away from the real axis. The same statement can be made even when there are real zeros so long as the step response corresponding to the real poles and real zeros alone is nondecreasing. Such behavior can be expected since indefinitely increasing the imaginary parts of the zeros moves them out toward infinity resulting in a system function whose step response is nondecreasing. This result may be established as follows:

Theorem 2: Let

$$W(s)=[(s+\alpha)^2+\beta^2]F(s) \tag{32}$$

where $\beta>0$, α has any real value, and the poles of the rational function $F(s)$ are all real. Also, let the poles of $F(s)$ outnumber its zeros by at least three. If the step response corresponding to $F(s)$ is nondecreasing, then the step response corresponding to $W(s)$ will be nondecreasing for β sufficiently large.

Proof: Replacing $s+\alpha$ by s corresponds to multiplying the impulse response by exp (αt). This will not change the sign of the impulse response at any point. Hence, it is sufficient to establish that the impulse response is nonnegative in the case where α equals zero. Such a shift in the pole and zero pattern may place real poles on the positive real axis but this is of no consequence in the following argument.

Denote the unit impulse response of $W(s)$ and $F(s)$ by $w(t)$ and $f(t)$, respectively. Since the poles of $F(s)$ outnumber its zeros by at least three, $f(0+)$ and $f^{(1)}(0+)$ are both zero. Hence, by equation 32, wherein $\alpha=0$,

$$w(t)=f^{(2)}(t)+\beta^2 f(t) \tag{33}$$

Now, $f(t)$ is a finite sum of terms of the form,

$$A_i t^{\mu_i-1}\, e^{-\rho_i t} \tag{34}$$

where A_i and ρ_i are real and μ_i is a positive integer. Since $f(t)$ is nonnegative, there is a value of time T beyond which $f^{(2)}(t)$ is positive, except possibly in the case where the real pole of $F(s)$ furthest to the right is at $s=0$ and is of single or double multiplicity. Hence, by equation 33, $w(t)$ is positive for $T\leq t<\infty$ and for all $\beta>0$. In the exceptional case the same conclusion follows since $f^{(2)}(t)$ becomes arbitrarily small with respect to $f(t)$ as t approaches infinity.

Furthermore, $f^{(1)}(t)$ is again a finite sum of terms of the form 34. This function is analytic for all finite t, considering t to be a complex variable. But the zeros of any analytic function are isolated[21] and hence $f^{(1)}(t)$ can have only a finite number of zeros on the real interval $0\leq t\leq T$. Therefore, $f(t)$ has only a finite number of maxima and minima for $t\geq 0$ and there are only a finite number of points at which $f(t)$ is zero.

In the neighborhood of each minimum of $f(t)$, $f^{(2)}(t)$ is nonnegative. Since $f(t)$ is also nonnegative, a positive number A, sufficiently small, can be found such that, for all intervals within which $0\leq f(t)<A$, $f^{(2)}(t)$ will be nonnegative. Denote this finite set of intervals by I. By equation 33, $w(t)\geq 0$ for t in I.

Moreover, $f^{(2)}(t)$ is a bounded function in $0\geq t\geq T$. Let $-B$ $(B>0)$ be a lower bound on $f^{(2)}(t)$. In the complement of I with respect to $0\leq t\leq T$, $f(t)\geq A>0$ and $f^{(2)}(t)>-B$. So by equation 33, $w(t)\geq -B+\beta^2 A$. Thus, for $\beta^2>B/A$, $w(t)$ will be nonnegative in the complement of I and hence for all t. This completes the proof.

An immediate consequence of theorem 2 is that, starting with a nondecreasing step response, we may generate another nondecreasing step response by adding any number of pairs of complex zeros so long as the resulting system function has more poles than zeros and the imaginary parts of the added zeros are large enough.

It should be noted that the result given by theorem 2 can be stated in the case where $F(s)$ is allowed to have complex poles if one is not interested in the behavior of the step response after it remains within some given deviation ϵ from its final value. This is the situation in any practical instance. In this case a time T can be found beyond which $|a(t)-a(\infty)|<\epsilon$. In the time interval, $0\leq t\leq T$, $f(t)$ will have only a finite number of points where it is zero and the preceding proof for the interval, $0\leq t\leq T$, applies in exactly the same way. This may be stated as follows:

Corollary 2a: Let

$$W(s)=[(s+\alpha)^2+\beta^2]F(s)$$

where $\beta>0$, α has any real value, and the poles of the rational function $F(s)$ outnumber its zeros by at least three. Also, let the step response corresponding to $F(s)$ be nondecreasing in the interval $0\leq t\leq T$. Then, the step response corresponding to $W(s)$ can be made nondecreasing in the same interval by choosing β sufficiently large.

The Lukacs-Szasz theorem 3 serves as an example of this corollary. In this case T may assume any finite value.

Appendix

Elgerd and Stephens[22] have computed a large number of transient responses corresponding to a variety of pole and zero patterns. Some of these serve as examples for the foregoing results.

References

1. Network Realizability in the Time Domain, **A. H. Zemanian.** *Transactions*, Professional Group on Circuit Theory, Institute of Radio Engineers, New York, N. Y., vol. PGCT-6, Sept. 1959, pp. 288–91.

2. Uber Positive Fourier-Integrale, **M. Mathias.** *Mathematische Zeitschrift*, Berlin, Germany, vol. 16, Feb. 1923, pp. 103–25.

3. The Laplace Transform (book), **D. V. Widder.** Princeton University Press, Princeton, N. J., 1946.

4. Monotone Funktionen, Stieltjessche Integrale und Harmonische Analyse, **S. Bochner.** *Mathematische Annalen*, Leipzig, Germany, vol. 18, 1933, pp. 378–410.

5. On the Design of Four-Terminal Interstages for Pulse Applications, **A. K. Choudhury, N. B. Chakrabarti.** *Indian Journal of Physics*, Calcutta, India, vol. 40, Apr. 1957, pp. 193–210.

6. A Necessary and Sufficient Condition for a Bounded, Nondecreasing Step Response, **M. R. Aaron, R. G. Segers.** *Transactions*, Professional Group on Circuit Theory, Institute of Radio Engineers, vol. PGCT-5, Sept. 1958, pp. 226–27.

7. On the Representation of a Function by Certain Fourier Integrals, **H. Cramer.** *Transactions*, American Mathematical Society, New York, N. Y., vol. 46, Sept. 1939, pp. 191–201.

8. Zur Kennzeichnung der Characteristischen Funktionen, **A. Khintchine.** *Bulletin*, State University of Moscow, Moscow, USSR, vol. 1, 1937, pp. 1–3.

9. Transient Response Conditions for Monotonic Characteristics, **I. F. Macdiarmid.** *Wireless Engineer*, London, England, vol. 28, Nov. 1951, pp. 330–34.

10. Monotonic Transient Response, **O. P. D. Cutteridge.** *Proceedings*, Institution of Electrical Engineers, London, England, vol. 101, pt. IV, 1954, pp. 46–54.

11. The Effect of Pole and Zero Locations on the Transient Response of Linear Dynamic Systems, **J. H. Mulligan, Jr.** *Ibid.*, Institute of Radio Engineers, vol. 37, May 1949, pp. 516–29.

12. Certain Fourier Transforms of Distributions—II, **E. Lukacs, O. Szasz.** *Canadian Journal of Mathematics*, vol. 6, 1954, pp. 186–89.

13. Certain Fourier Transforms of Distributions, **K. Takano.** *Tohoku Mathematical Journal*, Sendai, Japan, vol. 3, 1951, pp. 306–15.

14. The Inversion of the Laplace Integral and the Related Moment Problem, **D. V. Widder.** *Transactions*, American Mathematical Society, vol. 36, Jan. 1934, pp. 107–200.

15. Conditions on Pole and Zero Locations Which Insure a Nondecreasing Step Response, **A. H. Zemanian.** *Transactions*, Professional Group on Circuit Theory, Institute of Radio Engineers, vol. PGCT-6, Mar. 1959, p. 130.

16. On the Pole and Zero Locations of Rational Laplace Transformations of Nonnegative Functions, **A. H. Zemanian.** *Proceedings*, American Mathematical Society, vol. 10, Dec. 1959, pp. 868–72.

17. Some Nonnegative Trigonometric Polynomials Connected with a Problem in Probability, **E. Lukacs, O. Szasz.** *Journal of Research*, National Bureau of Standards, Washington, D. C., vol. 48, Feb. 1952, pp. 139–46.

18. Nonnegative Trigonometric Polynomials and Certain Rational Characteristic Functions, **E. Lukacs, O. Szasz.** *Ibid.*, vol. 52, Mar. 1954, pp. 153–60.

19. Uber Kosinuspolynome und die Nullstellen von Polynomen, **J. F. Schultze.** *Jaresbericht Deutscha Mathematische Verein*, Berlin, Germany, vol. 50, 1940, pp. 35–40.

20. Time-Response Characteristics of a System as Determined by its Transfer Function, **J. D. Brule.** *Transactions*, Professional Group on Circuit Theory, Institute of Radio Engineers, vol. PGCT-6, June 1959, pp. 163–70.

21. The Theory of Functions (book), **E. C. Titchmarsh.** Oxford University Press, London, England, second edition, 1939, p. 88.

22. Effect of Closed-Loop Transfer Function Pole and Zero Locations on the Transient Response of Linear Control Systems, **O. I. Elgerd, W. C. Stephens.** *AIEE Transactions*, pt. II (*Applications and Industry*), vol. 78, May 1959, pp. 121–27.

Static Control for a Mechanically Regulated D-C Supply

H. J. ABRAMS
NONMEMBER AIEE

J. F. BRUBAKER
MEMBER AIEE

A NEW APPROACH to automatic voltage control of a regulating transformer was sought in conjunction with a rectifier transformer and semiconductor rectifying elements connected in a 6-phase double-way circuit. Conventional control for a load tap changer is not designed to detect and correct for fluctuations directly in a d-c line, yet close control of the voltage across the 40-kv rectifier output was required for a radar pulsing station. The alternative of a-c sensing at a point just ahead of the rectifier transformer was considered but this was abandoned due to the problems of reactive drop compensation.

This paper is intended to lay the background for and to describe the method by which the problem was solved with the design of a new d-c sensing control, using two high-gain transistor-magnetic bistable amplifiers and a varistor-detector. Power gains of 10^8 with 20-millisecond response time and a drift level of 10^{-8} watts in the bistable amplifier reduced the system accuracy to that of the varistor-detector. Nevertheless, over an ambient temperature range of 25 to 55 degrees centigrade the trip-point accuracy of the system can be held to better than 0.1%. The success of the control may set the trend for other applications where the output to be controlled is direct current. The low-voltage tap on a bleeder resistance bank across the 40-kv line is the link that made the application of this control possible. There is no foreseeable difficulty in controlling voltages up to 300 kv by the same technique.

Over-all System Operation

Function of Regulating Transformer

To describe clearly the nature of the over-all problem it would be well first to review the function of regulating transformers as a class. In power apparatus a regulating transformer (synonomous with power regulator) is distinguished from the familiar step-up or step-down power transformer in that its ratio is not fixed but rather is variable within certain limits, usually ±10%. Its kilovolt-ampere parts are determined in general by its range of regulation relative to the kilovolt-amperes of the system. Regulation is made possible by one or more automatically controlled 32-step load tap changers interconnected with the windings of the regulating transformer.

In the case under discussion the range of regulation is extremely high by normal standards, inasmuch as the output voltage must be variable within limits of zero and 100% of the 4,160-volt 3-phase sup-

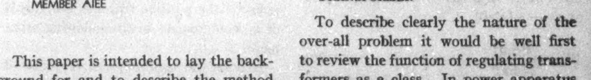


Fig. 1. Block diagram of 40-kv d-c supply system

Paper 60-783, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting, Atlantic City, N. J., June 19–24, 1960. Manuscript submitted November 24, 1959; made available for printing March 29, 1960.

H. J. Abrams is with the Norbatrol Electronics Corporation, Pittsburgh, Pa., and J. F. Brubaker is with the Westinghouse Electric Corporation, Sharon, Pa.

The authors wish to acknowledge the assistance of A. Maruschak, G. D. Throop, D. K. Barnes, and A. W. Possner.

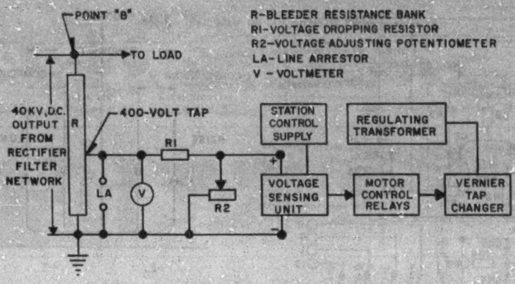

POINT "B"
TO LOAD
40KV.D.C. OUTPUT FROM RECTIFIER FILTER NETWORK
R
400-VOLT TAP
LA
V
R1
R2
+
−
STATION CONTROL SUPPLY
REGULATING TRANSFORMER
VOLTAGE SENSING UNIT
MOTOR CONTROL RELAYS
VERNIER TAP CHANGER
R-BLEEDER RESISTANCE BANK
R1-VOLTAGE DROPPING RESISTOR
R2-VOLTAGE ADJUSTING POTENTIOMETER
LA-LINE ARRESTOR
V - VOLTMETER

Fig. 2. Block diagram of voltage-regulating control system

ply system voltage, with an additional extension to take care of a possible ±5.0% variation in the supply. In Fig. 1, a simplified block diagram of the basic system components, the regulating transformer may be thought of as representing three separate cores housed within a single oil-filled case. These cores and associated windings are identified by function as: 1. the exciting autotransformer, 2. the series transformer for the coarse tap changer (so named because of the relatively large 2.5% voltage changes per step), and 3. the series transformer for the vernier tap changer (so named because of the relatively small 1.0% voltage changes per step). The output of the regulating transformer is applied to a step-up rectifier transformer capable of maintaining a maximum of 40,000 volts at 840 kw on the d-c side of a rectifier.

FUNCTION OF AUTOMATIC CONTROL EQUIPMENT

It is the function of the automatic control equipment to provide intelligence for correct operation of the load tap changers in accordance with a predetermined operational sequence, and subject to load demands. The special nature of this particular system's operation requires that there be three stages of automatic control, each initiated directly or indirectly by external events.

1. *Start-Up Stage.* Load voltage is raised from a predetermined minimum to the approximate operating level in about 25 seconds of elapsed time after closing of the load breaker. The coarse tap changer automatically stops when the preset manually selected tap has been reached.

2. *Operating Stage.* The control transfers to an automatic voltage-regulating unit which employs the vernier tap changer to maintain finer control of output voltage and to compensate for fluctuations in the supply. The voltage level to be maintained is preselected manually for the control unit.

3. *Shut-Down Stage.* Upon tripping of the load breaker, control shifts once more to the coarse tap changer and lowers the voltage to the predetermined minimum value. It is then ready for possible recycling.

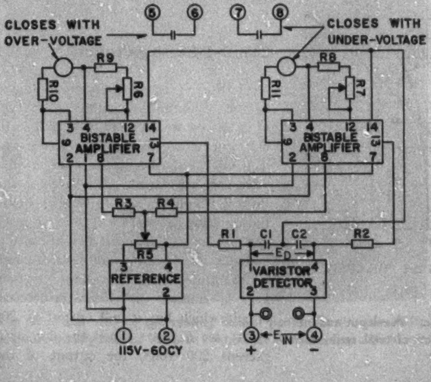

CLOSES WITH OVER-VOLTAGE
CLOSES WITH UNDER-VOLTAGE
⑤
⑥
⑦
⑧
R9
R10
R6
R8
R11
R7
BISTABLE AMPLIFIER
BISTABLE AMPLIFIER
3
4
12
14
9
13
2
1
8
7
R3
R4
R5
REFERENCE
R1
C1
C2
R2
E_D
VARISTOR DETECTOR
①
②
115V-60CY
③
④
E_IN
+
−

Fig. 3. Schematic diagram of d-c sensing control

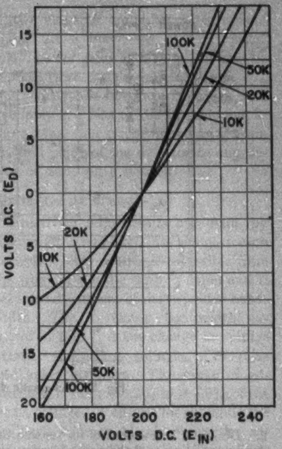


Fig. 4. Transfer curves for varistor-detector

VOLTAGE CONTROL DURING OPERATING STAGE

This paper concerns itself primarily with the operating stage as described in item 2, during which the control must continually measure changes in load voltage, and correct for them automatically within the range of the vernier tap changer. In power regulator applications voltage is seldom detected at the true load center. An accurate image of the load-center voltage is available at the output of the regulator through the use of a compensator to simulate line drop at any given load. Thus, if voltage must be maintained at point *B* but detected at point *A* in Fig. 1, a compensator is required with values of resistance and reactance chosen so as to represent the resistance and commutating reactance associated with the d-c load and the rectifier transformer. This is the conventional solution to the problem and, for the ideal case of a given number of rectifiers fed from a common supply system, it would fall within the capabilities of conventional a-c voltage-regulating relays and compensators.[1]

Choosing Voltage-Sensing Method

Consider two alternatives facing the designer at an early point in the planning of the voltage control:

1. Employment of conventional a-c sensing at point *A* in Fig. 1 with a compensating device to provide an a-c image of the d-c output at point *B*.

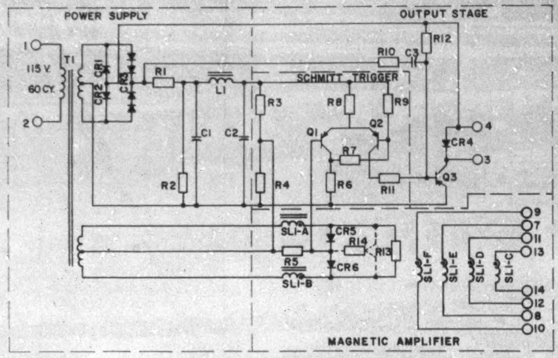


Fig. 5. Schematic diagram of bistable amplifier

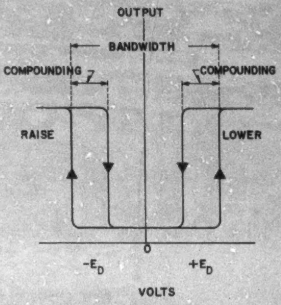


Fig. 7. Bandwidth and compounding characteristics of control system

2. Design of a new control for sensing the d-c output at point *B* through a bleeder tap.

Indirect Sensing of D-C Output

Two advantages stand out immediately in favor of alternative 1. The a-c control is standard and available. Furthermore, complete isolation from the line voltage is afforded by the connection of a potential transformer at point *A*. An obstacle looms, however, in the choosing of compensation constants to simulate the reactive drop of the circuit. This arises from the certainty that in system operation a varying number of in-phase rectifiers will be fed from the common a-c source. Thus the effective commutating reactance[2] to be considered is not a predictable constant but is instead dependent on the number of rectifiers being supplied simultaneously through the common source reactance. With alternative 1 it is therefore not possible to produce an accurate image of the d-c output for an individual rectifier at all possible values of the total system load.

Direct Sensing of D-C Output

The choice of alternative 2 eliminates the need for compensation, thus focusing attention entirely on the development of a control sensitive to average voltage deviations in the 40-kv d-c output.

Fig. 2 illustrates in block form the arrangement of components used to provide this type of control. Advantage is taken of the availability of a bleeder-resistance bank *R* in shunt with the 40-kv line and having a 1.0% tap for metering purposes. An arrester is connected between this 400-volt tap and ground to protect the sensing circuit from accidental overvoltages. Remote switchboard mounted voltmeter *V* and potentiometer *R2* facilitate the selection of the desired output voltage within a specified range. Resistor *R1* limits the voltage obtainable across *R2* to a value well within the limitations of the sensing unit.

There is no output from the voltage-sensing unit when the input is exactly 200 volts. A signal is delivered to the motor-control relays of the vernier tap changer only when the deviation is greater than the bias setting of either the raise or the lower bistable amplifier. The effect of tapping in more resistance at *R2* is to maintain a lower voltage level on the main rectifier output. Conversely, tapping out more resistance at *R2* has the effect of maintaining a higher output level.

The D-C Sensing Control

The d-c sensing control system is similar to the well-known balanced beam type of relay control, except that all static components are used. Attempts to use as many standard components as possible were very successful. The detector and bistable amplifiers are off-the-shelf items, but coupling networks had to be designed and a few adjustment potentiometers included for the compounding and bandwidth adjustments. The basic system, Fig. 3, consists of a varistor-detector circuit which has a null output at 200 volts; see Fig. 4. When the d-c input exceeds 200 volts, the output of the

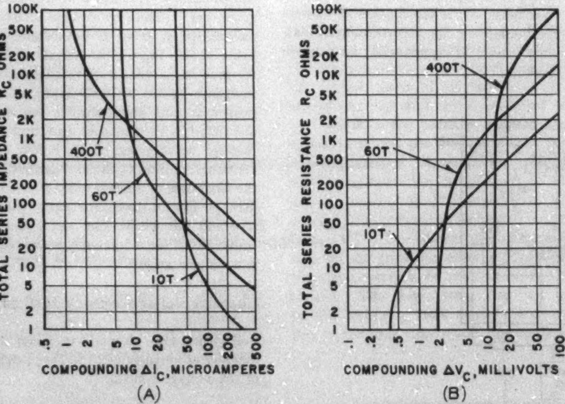


Fig. 6. Input current and voltage required to actuate bistable amplifier. A—Input current as a function of total series impedance. B—Input voltage as a function of total series resistance

Vol.
Pt. 1
196

Table I. Test Data on D-C Sensing Panel

Station Supply		Sensitivity	Volts E_{in}			
			Undervoltage Relay		Overvoltage Relay,	
Voltage	Frequency, Cycles per Second	Potentiometer R5 Setting*	Close	Open	Open	Close
126	57	4	202.2	202.4	208.4	208.6
120	60	4	202.2	202.4	208.4	208.6
114	63	4	202.3	202.5	208.4	208.6
126	57	6	203.5	203.8	207.3	207.6
120	60	6	203.5	203.8	207.4	207.6
114	63	6	203.5	203.8	207.3	207.6
126	57	8	204.6	204.8	206.3	206.6
120	60	8	204.6	204.8	206.3	206.6
114	63	8	204.6	204.8	206.3	206.6

* See Fig. 3.

detector will be a positive voltage. Two bistable amplifiers are biased so that an adjustable dead zone is located around the zero axis. The bistables offer a 20,000-ohm load to the detector. With this load the gain of the detector is about 0.4 volt per volt. Thus, if the input voltage should increase 4 volts, a 1.6-volt signal will be fed into the bistable. This is more than sufficient to actuate the lower bistable, which switches from the off to the on state. This in turn causes the vernier tap changer to function and to lower the voltage as many steps as necessary to return the high voltage to the desired output value. If the input voltage is lower than 200 volts, the raise bistable will be actuated in a similar manner.

A varistor-detector with a relatively high operating voltage for null output was chosen. Its null point is temperature-compensated.

The bistable amplifiers, Fig. 5, in this control combine the advantages of both magnetic amplifier and transistor circuitry.[3] The input stage of the bistable amplifier is a half-wave push-pull magnetic amplifier. This, in turn, drives a transistor Schmitt trigger.[4] A Schmitt trigger is a flip-flop circuit which has two stable states, either on or off. The trigger may be adjusted so that with zero control it may be either at full output or zero output. An output transistor power stage, which is switched by the Schmitt trigger, delivers 24 volts at 210 milliamperes full-wave rectified direct current. Since the first stage is a half-wave push-pull magnetic amplifier, low drift and fast responses are assured. The output of the magnetic amplifier stage around null is a spike of positive or negative voltage. This voltage spike is used to trigger the transistor circuit into the on or off steady-state condition. A very small amount of power is necessary to change the spike from positive to negative, since the change in the average power output of the magnetic amplifier is extremely small. At a maximum of $1^1/_2$ cycles' delay, power gains as high as 10^3 may be obtained with this device; see Fig. 6.

A zener-diode reference circuit is used to supply the bias to the bistable amplifiers so that they can be biased away from the null point of the detector to any degree of bandwidth desired.

CONTROL ADJUSTMENTS

Sensitivity adjustments are made by changing the regulated bias signal to each bistable. This control is accessible from the front of the d-c sensing panel and it permits adjustment of the regulator bandwidth. The regulator bandwidth, as shown in Fig. 7, is the change in voltage between the pickup voltages, of the raise and lower bistables. The range of adjustment is from ±0.2% to ±5.0% of the input voltage. Controls for changing the feedback to each bistable amplifier are also available at the panel. This allows the compounding to be changed from 0.1% to 5.0% of the operating voltage. Compounding is defined here as the change in voltage E_d necessary to move from the on to the off state of either bistable amplifier.

PERFORMANCE

The performance just described can be realized with a ±5.0% change in the voltage and a ±3-cycle change in the frequency of the station control supply. Since the control will be located in a temperature-controlled environment, the ambient temperature will be held between 20 and 55 degrees centigrade. There is

Fig. 8. Front view of d-c sensing control panel, showing sensitivity control centrally located between controls for compounding adjustments

Fig. 9. Rear view of d-c sensing control panel, showing zener-diode reference, varistor-detector, and output relay cases mounted between the bistable amplifiers

essentially no change in performance over this temperature range. The complete control was not tested over a wide ambient temperature range, but the detector and the bistable amplifiers have been shown capable of operating over a much wider temperature range with very low drift levels.

A sample design will show that the null drift of the detector circuit and the percent voltage change per tap change are the limiting factors of the system accuracy. With a 200-volt E_{in} input to the detector, the detector's output voltage E_d will be zero. From Fig. 4 it may be seen that with a 20,000-ohm load, the detector's gain will be 0.4 volt per volt. If a 0.1% system were desired, it would be necessary to actuate the raise or lower voltage bistable with a ±0.2-volt change in E_{in}. With a detector gain of 0.4 volt per volt, the change in E_d for a 0.2-volt change in E_{in} is 0.08 volt.

From Fig. 6(B) it may be seen that, with a 400-turn control winding and a 20,000-ohm input impedance, it is possible to have compounding equal to 30 millivolts. The compounding power P_c, or the power necessary to move from the on to the off state is

$$P_c=\frac{(0.03)^2}{20{,}000}=4.5\times10^{-8}\text{ watts}$$

If the on point of each bistable is biased so that it takes a 0.08-volt change in E_d to turn it on, it would be possible to achieve a system with ±0.1% bandwidth and 0.038% compounding.

The foregoing analysis, of course, does not take into account the drift in the detector or that the value of the voltage change per tap is 1.0%. The use of such a low trip-point accuracy in conjunction with a tap changer having 1.0% steps would result only in unnecessary operations, known as tap-changer hunting. Consequently, the bandwidth setting recommended for this application is ±1.5 volts E_{in}, giving a system accuracy of 0.75%. Table I shows actual data on this system, tested with various line voltages and frequencies. The null characteristic of this particular varistor-detector varies somewhat from that shown in the typical curve data of Fig. 4.

The simplicity and compactness of the control is evident from an examination of Figs. 8 and 9. The panel assembly measures 10$^1/_2$ inches high, 11 inches wide, and 7 inches deep, thus comparing favorably in size with corresponding a-c sensing controls.

Conclusions

From standard static components a workable d-c sensing control has been developed that shows good promise. Its basic element is the bistable amplifier, acting as a sensitive relay in an on-off system for raise and lower voltage control. A varistor-detector, operated from a bleeder-resistance bank, senses changes in the output of a d-c power supply to trigger the amplifier. Fast response, widely adjustable bandwidth and compounding, and a low drift level in respect to varying ambient temperature and supply voltage and frequency, make it particularly suitable when it is used in conjunction with a regulating transformer feeding the rectifier-filter supply system for a radar pulsing station.

In addition to radar, this control is suitable to a wide field of other rectifier applications, including sonar, as well as equipment for proton acceleration. Rectifier systems presently deal in d-c outputs from 10 to 300 kv and in power levels as high as 3,000 kw. The control is not limited to regulators of the step type but may be used in a wide variety of applications where the output to be controlled is direct current.

References

1. **MAGNETIC AMPLIFIER CONTROL FOR LOAD-TAP-CHANGING EQUIPMENT, Bryan Webb, Jr., G. D. Throop.** *AIEE Transactions*, **pt. III (*Power Apparatus and Systems*), vol. 76, Dec. 1957, pp. 1085–90.**

2. **RECTIFIER TRANSFORMER CHARACTERISTICS, AIEE Committee Report.** *Ibid.*, **vol. 68, pt. II, 1949, pp. 1305–11.**

3. **A SENSITIVE STATIC RELAY CIRCUIT, John Rosa.** *Proceedings*, **Special Technical Conference on Nonlinear Magnetics and Magnetic Amplifiers,"** *AIEE Special Publication T-116*, **Sept. 1959, pp. 154–66.**

4. **A THERMIONIC TRIGGER, O. H. Schmitt.** *Journal of Scientific Instruments*, **London, England. vol. 15, Jan. 1938, pp. 24–26.**

Data Processing as a Tool for Generalizing Communications Systems

WILLIAM F. LUEBBERT
NONMEMBER AIEE

AUTOMATIC data processing is a tool for transforming, processing, and calculating information. If one accepts this fact, one should be prepared to describe the fundamental capabilities of automatic data processing. Since communications is itself a tool, it is instructive to attempt to describe its fundamental nature in a similar frame of reference and compare the two. The results of such a simple analysis lead to a viewpoint which has important implications on system and equipment design.

The fundamental capability of data-processing equipment can be described as the ability to transform information into more desirable or useful forms. From the same viewpoint, the fundamental capability of conventional communications equipment can be described as the ability to transfer information to a more desirable or useful location. Of course, from an abstract-logical viewpoint a transfer is merely one special kind of transformation, one in which only physical location is changed. Combining these descriptions, one is led to the concept of a generalized data-handling system capable of performing generalized transformations, of which conventional one-location data processing would be one special case and conventional no-processing data transmission would be another special case. In such a system there would be no fundamental distinctions between data processing and data transmission, but only distinctions of convenience based upon application, use, and design emphasis.

An examination of the information flow and manipulation in typical data-processing and communication equipments shows almost immediately that there is a very considerable mixture of functions going on in both types of equipment. A considerable amount of the activity going on inside any computer or data processor consists of simple transfers of information from one part of the processor to another.

At the same time within communications equipment a major part of the activity concerns generation, manipulation, and other processing of information used for control, supervisory, and error reduction purposes. In many cases nearly identical operations go on in both data-processing and communications equipments, with the differences, if any, being matters of design emphasis based upon application and use.

Is such a concept a reasonable one, and does it have any practical utility? The answer to the first part of this question will be examined in detail in this paper. Whether this concept has practical utility will be determined by the success in actual use of the equipment and system concepts it generates, which are receiving initial implementation in the FIELDATA family of equipments. These concepts are explained in this paper.

It is conventional to consider communications systems as consisting of three types of subsystems: subscriber, central-office, and transmission subsystems. The impact of automatic data-processing equipment, techniques, and concepts upon each of these three kind of subsystems will be illustrated by discussions of three topics:

1. The impact of data processors at subscriber locations, with particular emphasis upon the effects of processors upon the amount of communications required to meet user demands.
2. The impact of data processors on central office activities such as circuit of message switching.
3. The impact of data processing upon the design of data terminals which include transmission capabilities.

Impact of Data Processors at Subscribers' Locations

Experience to date shows that the introduction of data-processing equipment is often accompanied by an explosive increase in communications requirements. In part, this is due to the voracious appetite of data processors for current, up-to-date information in applications which take advantage of their abilities to process vast amounts of information economically. In part, it is because their tremendous speed of processing compared to the speed of human processors dramatically changes time factors in an over-all data-handling cycle.

For example, if a given report prepared by nonautomated techniques requires 45 days for its preparation, a delay of 3 or 4 days consumed in sending the raw data by mail is hardly noticed. However, if a data processor requires only one day to prepare the report from the raw data, this same 3- or 4-day delay in the mail becomes the overwhelming factor determining over-all report preparation time, and an obvious bottleneck. Clearly if the report is an important one, badly needed for effective planning, there is a strong tendency to demand better communications service to eliminate the bottleneck. Hence, considerable amounts of data once transferred by mail are frequency added to the electrical communications requirements.

This mushrooming of communications requirements is not, of course, an inevitable result of the adoption of data processing. On the contrary, one of the most attractive features of data processors to the communicator is their ability to process and reduce raw data to a more concentrated form, boiling out those portions which are irrelevant, redundant, or insignificant, so that the communications system and its human users need not be burdened. If a data processor is used to reduce data to one tenth of its raw form, it may then be used to reduce communications required to transmit that data by the same factor. Thus, data processors acting as data reduction devices located at an entry point of data into a communications system may be a particularly valuable tool for preventing large sources of raw data from overloading a communications system with excessive demands for service.

Of course, the same increased speed of processing by machines which disrupts the balance of many processing communications cycles, thus increasing the demand for electrical communications, may under other conditions have the opposite effect. If the key factor in a situation is the achievement of a fixed delivery deadline and the processing time is reduced, then the processing time saved becomes

---

Paper 59-1093, recommended by the AIEE Data Communication Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Chicago, Ill., October 11–16, 1959. Manuscript submitted May 11, 1959; made available for printing June 26, 1959.

WILLIAM F. LUEBBERT, formerly with the U.S. Army Signal Research and Development Laboratory, Fort Monmouth, N. J., is with the United States Military Academy, West Point, N. Y.

available for communications. For example, if a 15-day deadline exists and 14 of those days are required for manual processing of data, then only one day is available for transmittal of the information, and electrical transmission may be necessary to meet the deadline. However, if an automatic data processor can prepare the information in 2 days, the remaining 13 days allow ample time for transmittal of the information in more routine fashion by mail or messenger, thus reducing the demand for electrical communications.

Whether the factors tending to increase communications requirements or those tending to decrease them are dominant in a particular case when automatic data-processing is introduced is, of course, not so much a function of the data-processing equipment as it is of the situation, and the concepts and techniques of employment of the equipment.

Which of these factors become dominant in a particular case when data-processing equipment is introduced is not so much a function of the equipment used as it is of the concepts and techniques of employment used. The equipment will have an impact upon communications requirements, but it is only a tool, and the user and his method of employment of this new and powerful tool will, in the final analysis, determine what its impact will be.

Data Terminals

Many practical cases of the close interrelation of transmission and processing functions are immediately obvious in the area of devices used for data entry and output. For example, computers frequently use paper tape readers similar to those used in teletypewriter transmitter distributors, and paper tape punches that could easily be used in teletypewriter reperforators. Similarly, the idea that various kinds of input-output devices such as card readers and punches and magnetic tape transports, widely used for data processing, can effectively be adapted for the use of communications lines is also being exploited in equipments such as the IBM (International Business Machines Corporation) Transceiver, the Collins Kinetape, etc.

FIELDATA emphasizes this kind of exploitation to the extreme, particularly encouraging it by assuring that interconnection of semiautonomous equipment modules be made in accordance with common standards, regardless of whether these equipments are conceived primarily for computer-associated or transmission-associated functions. This makes it possible for the same data terminals to operate with not only the data-processing type of inputs and outputs (such as computers, paper tape, magnetic tape, and IBM cards) but also with real time weapons system data, and with telegraphic data.

Control circuitry is so devised that pure binary as well as alphanumeric (alphabetic-numeric) data may be handled. Since any digital code, be it Baudot (teletypewriter), Holerith (IBM card), or any of a wide variety of computer codes, may be represented in binary bit-by-bit form, the FIELDATA devices have the potential of transmitting or handling any type of digital data. Thus they could be used with digitized voice, digitized facsimile, or other types of digitized analog signals.

The use of common standards, codes, and standard data rates makes possible the kind of data transmission equipment concept which leads naturally to a division of the subassemblies of data transmission equipment into three kinds:

1. *Input-Output Transducers.* These are devices for converting information from some human- or machine-usable form such as paper tape, magnetic tape, punched cards, analog electrical voltages, strokes on a keyboard, etc., into digital form.

2. *Transmission Transducers.* These are devices for converting data in digital form into appropriate signals for transmission over radio, wire, or other kinds of propagation media.

3. *Embolic* Equipment.* This is equipment which is normally inserted between input-output transducers and transmission transducers; it is used primarily to perform control and supervisory functions, error detection and/or correction, buffering and/or speed conversion, code conversion, or encryption necessary for proper system operation of the data transmission equipment. The functions of embolic equipment are information-processing functions. Inputs and outputs will both be digital in form although supplementary analog information may also be available particularly in some kinds of error control schemes. A general purpose computer is potentially a very powerful and flexible type of embolic equipment, but the necessary functions can often be performed much more economically by specialized equipment.

This division of subassemblies may be a physical division into separate items of equipment, or into semiautonomous parts of a single equipment (either in separate boxes or in a single box), or it may be merely conceptual with no physical implementation.

Obviously, maximum flexibility and adaptability can be achieved if any input-output transducer can operate with any transmission transducer. This permits one to tailor an equipment to meet the requirements of a particular situation. It permits one to choose the in-out device suitable for the user's most convenient data form (paper tape, magnetic tape, IBM cards, etc.) independently of the nature of the transmission facility available. One can then choose the transmission transducer on the basis of the transmission medium used (very-high-frequency radio relay, high-frequency radio, loaded cable, carrier transmission on wire, etc.) independently of the user's data form. Then the two are joined together to advantage of common standards of interconnection or intercommunication between modules and to create a well-tailored combination.

The question then arises, "Why embolic equipment?" Certainly, by proper choice of common interconnection or intercommunications characteristics, one can minimize requirements for code conversion, buffering, speed conversion, etc. Ideally, one should be able to join the input-output transducer to the transmission transducer without embolic equipment, so why have it? The answer is that there are three important functions in a data terminal which seem convenient, to separate from both in-out transducers and transmission transducers: communications supervision, error control, and cryptographic security.

Supervisory functions are conveniently separated from the input-output transducers and the transmission transducers because supervisory requirements may be strongly influenced by both. Error control requirements may also be strongly influenced by both the in-out data and the transmission situation, and may even in some cases not be desired at all. Thus, it is best treated as a module which can be tailored to the paired requirements of in-out and transmission or completely omitted. Cryptographic security is conveniently a separate module so that it may be omitted in those cases where security is not required or desired.

Primary and Secondary Information

To discuss supervisory activities conveniently it is desirable to make a distinction between two kinds of information which flow through a communications system primary and secondary information.

---

* A coined term from the Greek *embolisimos*, to put between or insert. The word is also used in medical, astronomic, and ecclesiastic literature to describe other specific kinds of intercalations.

Primary information is that information which a user wishes transferred to another location, i.e., the information to be communicated.

Secondary information is that information added to the primary information either by the originator or by equipment of the communications system which is used to perform functions of supervision, routing, error control, and related activities necessary or desirable to permit the primary data to be effectively communicated. This information is used by communications equipment and personnel and is normally of no use or interest to the ultimate recipient of the primary information.

A basic requirement for maximum flexibility and adaptability is that the user, who enters data at an input device and receives data at an output device, need not be required to exercise judgment or knowledge, to perform special activities, or to use or interpret secondary information. This, in turn, makes it desirable to isolate secondary information from the input-output devices, which are the users' point of contact with the communications system. A convenient way of establishing and enforcing this isolation is the creation of a distinct embolic module which generates, receives, interprets, and/or acts upon secondary information and passes on action requirements derived from this secondary information over local control lines to its associated input-output transducer and transmission transducer. It is important to note that such an equipment need not have any ability to interpret or act upon primary data, and thus its operation can be made completely independent of the coding used for primary data so long as there is a unique method of distinguishing primary from secondary data.

In the FIELDATA family of equipments the segregation of primary and secondary data is accomplished in a particularly simple way by providing separate transmission symbol "alphabets" for primary and secondary data. The basic unit of data in FIELDATA codes is a 6-bit character. For primary data this may be merely a 6-bit block with no specific meaning assigned to specific bits within it, and no restrictions on the permissible code combinations which may be transmitted. If this information is alphanumeric in form so that specific bit meanings must be assigned for electric typewriters, printers, and similar input-output transducers to operate, then the FIELDATA alphanumeric code given in Appendix I is used. A similar but distinctly different 6-bit FIELDATA supervisory data code has also been created; see Appendix II. In situations where both FIELDATA primary data and FIELDATA supervisory code data might appear on the same six lines, the same six channels of tape, etc., a seventh bit is added to provide unique distinction. This seventh bit is normally a binary ONE for primary data and a binary ZERO for supervisory data. If, on a particular recording medium such as paper tape or on a particular kind of transmission channel, adequate reasons exist for tagging the difference between these two symbol alphabets in some other way, then other methods could be used. All secondary information except error control redundancy is handled by the FIELDATA supervisory data code in FIELDATA equipments.

The number, variability, and difficulty of measurement and analysis of the various factors which contribute to the frequency and interrelationships of transmission errors or particular circuits is staggering. At present, the state of the art is such that only crude estimates of frequency of error can be made for typical equipments when they are exposed to disturbances other than gaussian noise, and that practically nothing can be estimated in advance about the interrelationships of error under practical conditions of impulse noise, crosstalk, propagation variations, etc.

This is particularly unfortunate because the effectiveness of the various digital error control schemes available is strongly dependent upon the interrelationships of errors. For example, a simple parity check is capable of detecting single errors but not double errors. If errors occur randomly, double errors will seldom occur and this very simple check will be quite powerful. Thus, if the bit error rate is 10^{-4}, a simple parity check will reduce the undetected error rate to 10^{-8}. On the other hand, if errors tend to be clustered, a parity check will be rather ineffective. Thus, if the conditional probability of a second error immediately following the first is 0.5 and the bit error rate is 10^{-4}, then a parity check will reduce the undetected error rate to only 0.5×10^{-4}. Given knowledge of the interrelations, checks can be designed which give high protection with a minimum amount of checking equipment. Unfortunately this knowledge is usually unavailable.

If the most important sources of errors occur in a data communications link, is it proper to incorporate the major error control features of the link into the transmission transducers? The FIELDATA concepts answer this question with a very definite "No." The key reason for this is that while the *occurrence* of errors and the raw error rate and characteristics are determined primarily by transmission factors, the error *requirements* are determined by the use of the data. The means and techniques of error control appropriate to a particular situation obviously depend upon the interaction of these factors. However, it is a fundamental modularity principle of FIELDATA that the characteristics of transmission transducers should as nearly as possible be independent of the details of user input-output characteristics and data employment.

If one were to incorporate the error control features into the transmission transducers and one desired to provide p classes of controlled error service to users with q modulator-demodulator assemblies, then one would require p times q types of complete transmission transducers. The obvious answer is to separate modularly, making the error control module an item of embolic equipment. This allows one advantage of similarities in requirements and raw error characteristics among the different situations to reduce the variety of equipments to be constructed.

In view of the present difficulties in predetermining the specific error characteristics of transmission transducers prior to construction and test, it also permits construction of new transmission transducers and their use with existing error control embolic equipments until the specific error characteristics of the transducer can be measured, and new error control embolic equipment designed if necessary to meet user requirements with the measured transmission error characteristics.

In addition to these practical advantages of placing error control responsibilities in embolic equipment modules rather than in transmission transducer modules, there are conceptual advantages associated with maintaining the simplest possible information flow patterns and a division of activities among the three basic kinds of assemblies.

In general, transmission transducers pay no attention to the information content of the digital information they convert to modulated transmission form, neither knowing nor caring whether the data is primary or secondary data, whether it is redundant or irredundant, or what code or codes it uses. In contrast to this, embolic equipments normally act as information-processing devices. Thus, in supervising a transmission link, embolic supervisory equipments act on sensory information received from transmission transducers and in-out transducers. They generate, process, and interpret secondary supervisory information to control the over-all operation of the communications link.

The error control problem is exactly parallel. Acting on information about user requirements (from the in-out transducer side) and on information and sensory information about transmission errors from the transmission transducer side, error control equipment is required to generate, process, and interpret error control information, using it to control (often via supervisory operations) the over-all operation of the communications link in such a way as to control its errors. Thus it is obvious that from the information processing viewpoint the performance of error control as an embolic function has a close parallel to other embolic functions and is distinctly different from the functions otherwise performed by transmission transducers.

Among the Fieldata equipments being developed in implementation of these concepts are the following:

Transmission transducers

- *AN/TSQ-32;* 600–1,200 bps (bits per second); Stelma Corporation.
- *AN/TSQ-33;* 600–2,400 bps; Collins Radio Company.
- *AN/TSQ-35;* 19,200 bps; Bendix Aviation Corporation. Bendix Pacific Division, Bendix Aviation Corporation.

General-purpose computers

- Mobidic; Sylvania Electric Products, Inc.
- Basicpac-Logipac; Philco Corporation.
- Informer-Data Coordinator; International Business Machines Corporation

In-out transducers:

- Electric typewriter; Smith-Corona, Inc.
- Paper tape reader; Smith-Corona, Inc.
- Paper tape punch; Smith-Corona, Inc.
- High-speed printer; Anderson-Nichols and Company.
- Paper tape transport; Ampex Corporation.
- Magnetic tape transport; Ampex Corporation.
- Tacden; Aeronutronics, Division of Ford Motor Company

Special embolic equipment

- *CV-689;* cryptosecurity adaptor; Collins Radio Company.
- *CV-690;* control equipment; Collins Radio Company.
- *CV-691;* data concentrator; Collins Radio Company.

The transmission transducers of the Fieldata family are limited in number. However, the choice of the 75×2^x pattern of data rates permits widespread augmentation by minor modification of existing or developmental teletypewriter multiplexed transmission equipments. In addition, future expansion is simplified by the fact that future equipments will be able to utilize the same embolic and input-output equipments, and will thus be cheaper to develop than data transmission equipments which require development of embolic and/or input-output equipment as part of the same package. Expected future expansion will place greater emphasis on transducers for radio circuits.

The *AN/TSQ-32* is a transmission transducer capable of accepting data from a standard Fieldata connection and of transmitting it serially as frequency shift modulation of a single subcarrier over a standard voice channel. Transmission rates are 1,200 bps or 1,500 wpm (words per minute) over good quality circuits, and 600 bps over poorer circuits.

The *AN/TSQ-33* is a transmission transducer capable of accepting data from a standard Fieldata connection and of transmitting it as eight channels, as synchronous phase quadrature modulation of four subcarriers over a standard voice channel. Transmission rates are 2,400 bps (3,000 wpm) over good quality circuits, with 1,200 and 600 bps available for use over poorer circuits. This equipment is essentially a militarized, miniature compatible Fieldata version of the Kineplex *TE-206*.

The *AN/TSQ-35* is a transmission transducer capable of accepting data from a standard Fieldata connection and of transmitting it by amplitude modulation of eight subcarriers orthogonally spaced in frequency and time. The transmission base band is the 48-kc band between 12 and 60 kc used for military and commercial cable and carrier circuits. Representative circuits are the Army spiral-4 and associated *AN/TCC-7*,- 8, and -11 equipments; and commercial types *N* and *K* carrier equipments. In addition to point-to-point full duplex operation this equipment has special features for the multiple-station "common net" round-robin type of operation.

In addition to these transmission transducers specifically designed for Fieldata a wide variety of existing military and commercial teletypewriter transmission equipments can be used for Fieldata transmission with only minor modification. Examples of such equipment are the *AN/FGC-54* capable of transmission and diversity reception of Fieldata information at 2,400 bps over long-haul 3-kc radio circuits using 32 channels, which operate at 75 bps. Another example is the *AN/FGC-29*, potentially capable of transmission and diversity reception Fieldata information at 1,200 bps using 16 channels which operate at 75 bps. Yet another example is the *AN/TCC-30* potentially capable of transmitting 1,200 bps using 16 channels which operate at 75 bps.

The Fieldata computers form the most complete and well-balanced portion of the Fieldata family. They may act as either input-output transducers or as embolic equipments for data transmission purposes, having the ability to accept, process, and emit either primary or secondary data. All are designed for direct operation with data transmission equipment. In their employment the Mobidic, which was the first machine to have its characteristics "frozen," is the least capable machine with regard to data transmission having serious restrictions on its use of the supervisory code functions. The most flexible in its employment of data transmission will be the data co-ordinator, a newer equipment which will be capable of terminating a large number of data transmission circuits simultaneously, and which will have a number of special capabilities and console positions to facilitate its use as a facilities co-ordination processor for an integrated communications data-processing system.

All the Fieldata computers are general purpose processors of modular design and great flexibility. All of these computers are designed for field use with operation and maintenance simple enough for field personnel. The largest, Mobidic mounts in a semitrailer van, while the others mount in shelters which can be carried on a truck.

The input-output transducers of the Fieldata family serve double duty as computer input-output devices and as data transmission input-output devices. The group under current development constitutes a minimum group of general usage items, a number of which are in only partially militarized form. This minimum group will be augmented by future field teletypewriter equipments which will utilize Fieldata code, and by advanced equipments now under study to provide specialized input-output capabilities. The items are mostly quite conventional, and detailed descriptions are omitted here.

The specialized embolic equipments in Fieldata provide cryptographic security, interconnection of input-output and transmission transducers, and tie-in of Fieldata circuits to existing teletypewriter circuits.

Three major items of special embolic equipment are being developed for FIELDATA. The *CV-689* is a special cryptosecurity adaptor which permits an existing type of security equipment to be inserted just before the transmission transducer in any FIELDATA data transmission assembly, thus providing cryptographic security.

The *CV-690* is a device which provides supervisory control, error control, and synchronizing buffer facilities for connecting paper tape or magnetic tape units to the *AN/TSQ-32* or *-33*. Although future plans call for similar special militarized embolic equipments for other kinds of input-output equipment such as card equipment, none is now under development. However, rather minor modifications of commercial Collins Kinecard equipment will permit nontactical employment of *AN/TSQ-33* equipment for card transmission, and at least some versions of the MOBIDIC will include card equipment which, through the computer, can reach transmission facilities.

Although FIELDATA concepts make provision for a wide variety of error control and supervisory control systems, only the particular system used in the *CV-690* will actually be used initially, except for computer-to-computer transmission, since other embolic devices using different error control or supervisory control schemes are now under development. When experiments do determine the actual frequency and interrelationship of errors for particular transmission transducers, it is expected that more effective schemes will be devised. However, the very simple 2-dimensional (interlaced) parity error detection scheme followed by the request-back or rerun type of error correction used in the *CV-690* is particularly easy to implement, and is expected to suffice for initial testing.

The *CV-691* data concentrator is a device designed to bridge the gap between existing large-scale 60- and 100-wpm teletypewriter facilities and FIELDATA equipment. Although all FIELDATA processors have normal provision for a paper tape reader which accepts teletypewriter tape, as well as one for FIELDATA tape, there exists a significant need, especially at locations where FIELDATA processors might not be available, to accept multiple channels of teletypewriter information and convert this into FIELDATA form to take advantage of FIELDATA transmission transducers and error control, permit recording on FIELDATA magnetic or paper tape, permit printing on FIELDATA high-speed printers, simplify entry into FIELDATA computers, etc. The *CV-691* accepts up to 25 (or 50) teletypewriter inputs, stores the information in a buffer core memory (made up of the same memory planes as used in MOBIDIC), assembles it into message blocks, converts it into FIELDATA form, applies the same error and supervisory control as the *CV-690*, and emits the data in FIELDATA form at rates up to 2,400 bps (3,000 wpm). The receive side performs the inverse functions.

As more and more of the voice and other analog communications systems convert to pulse code modulation and other digital forms, it is expected that additional types of FIELDATA embolic equipment will be required to perform the error control, supervisory functions, and buffering synchronization necessary to tie input-output transducers to their digital bit streams.

The FIELDATA family is an attempt to create an integrated family of data transmission equipments to meet Army needs. Though it lacks many of the features and equipments of an ideal family of data transmission equipments, it will make available in experimental quantities by the end of 1960 the first integrated family of equipments for experimental establishment of a truly integrated communications data-processing system. It is the first system in which both data-processing and communications equipment utilize the same input-output and storage devices, the same voltages, impedance levels codes, and other common interconnection characteristics; it is the first in which the equipments are so designed that in many cases the only way to determine whether a device is used for communications or data processing is to look at its specific application in the system.

Appendix I. Fieldata Alphanumeric Code

The following table represents the alphanumeric code for alphanumeric information in specific form.

	000	001	010	011	100	101	110	111
000	MSP	SET	RESET	LF	CR	SP	A	B
001	C	D	E	F	G	H	I	J
010	K	L	M	N	⊙	P	Q	R
011	S	T	U	V	W	X	Y	Z
100	)	–	+	<	=	>	–	$
101	*	(	"	:	?	!		⊕
110	0	1	2	3	4	5	6	7
111	8	9	'	;	/	.	SPEC	IDLE

MSP: Called "Master Space"; printed symbol ≷

SET: Causes a change of machine state or switch position. When translating to or from teletypewriters, this is interpreted as "Figures Shift"; printed symbol ζ

RESET: Causes the opposite change of machine state or switch position. When translating to or from teletypewriters, this is interpreted as "Letters Shift"; printed symbol ⁒.

CR: Carriage return; printed symbol <

LF: Line feed; printed symbol ←.

SPEC: Called "Special." Used as an escapement or alphabet expander. SPEC followed by another symbol forms a symbol pair with a new meaning which can be arbitrarily defined in any particular system. For example, SPEC A might be defined as α, SPEC B as β, etc. Extent of influence is only one symbol. However, the combination SPEC SPEC may be used to indicate that the characters following are to be interpreted by reference to a table of symbol assignments; printed symbol □.

IDLE: No information content; a deletable character.

Appendix II. Fieldata Control and Supervisory Code

Another 6-bit FIELDATA supervisory data code is described below. Though similar to that of Appendix I it has distinct differences.

Most Significant Bits	Least Significant Bits 000	001	010	011	100	101	110	111
000	←——Format effectors					——→	a	b
001	c	d	e	f	g	h	i	j
010	k	l	m	n	o	p	q	r
011	s	t	u	v	w	x	y	z
100 101 110 111	Various system supervisory and control symbols such as "start of block," "end of block," "end of file," "are you ready?," "I am ready," "please retransmit last block," etc.							

In FIELDATA standard form, items, in the alphanumeric code, or other user (primary) data, are designated with a control bit in the ONE condition. If the communicator desires to restrict himself to the bottom half of the control code, eschewing the alphabet made available in the top half, then the resultant code is self-clocking because an OR of the control bit and of the most significant bit of the code will always contain a ONE.

If in a particular system it is desired to provide 2-case alphabetic operation, i.e., *a* or *A*, *b* or *B*, etc., for the easy-read or business correspondence type of format, one may absorb the alphabetic portion of the control code into the user (primary) information code. This gives a user alphabet of 96 symbols and a control and supervisory alphabet of only 32. In such a case a 2-bit test is necessary to separate control functions: i.e., both the control bit and the most significant information bit must be tested. Treating the control bit as the least significant bit for sorting a binary sort will result in sorting in the order *a, A, b, B, c, C*, etc. By treating it as the most significant bit the two cases will be alphabetized separately, i.e., *a, b, c, d*, etc.; then *A, B, C, D*, etc.

Moon Reflection Studies with Bistatic Radar at 3,000 Mc

A. W. STRAITON
NONMEMBER AIEE

C. W. TOLBERT
NONMEMBER AIEE

MEASUREMENTS have been made of moon-reflected radar signals transmitted by the Royal Radar Establishment from Malvern, England, and received by the Electrical Engineering Research Laboratory of The University of Texas at Austin, Texas. These supplement monostatic radar measurements made by a number of organizations and a few bistatic measurements which have been noted in news articles and mentioned at technical symposia. Reports have been made of monostatic measurements in the very-high-frequency (vhf) band by Trexler and Evans and in the 3,000-mc (megacycle) region by Yaplee and Hughes.[1–4]

The measurements reported in this paper were made to demonstrate the feasibility of receiving moon-bounce signals approximately one fourth of the distance around the earth and to obtain information on the character of the signals received over this path. The signal study should shed light on the surface of the moon, distance to the moon, the Faraday rotation in the ionosphere, and the motion of the moon.

Theoretical Considerations

It is obvious that signal detection is made easier by an increase in transmitted power and in the gain of the transmitting and of the receiving antennas. For the higher frequencies, the gain for a given size antenna is increased but so also is the cost of construction.

Bandwidth

The signal-to-noise ratio may be increased by decreasing the bandwidth. This bandwidth must, however, be wide enough to accept any changes in frequency due to Doppler effects associated with the earth's rotation, the moon's libration, and changes in distance between the earth and the moon. The bandwidth must also be wide enough to avoid pulse distortion. For this latter reason, it is possible to make measurements with comparably low-power continuous wave operation, while higher peak power is needed for pulse operation.

Scattering Cross Section of the Moon

The only uncertain factor in the radar equation, as applied to the moon returns, is the back-scattering cross section of the moon. This reflection is dependent on the smoothness of the moon surface relative to the wavelength used. For wavelengths of several meters, the moon's surface has been described by Siegel as quasi-smooth, implying that large areas of the moon are relatively smooth for this wavelength.[5] It would be expected, however, that for centimeter wavelengths, the moon's surface would appear rougher and have more areas which would serve as effective reflecting regions.

In addition to the roughness, the reflections from the moon will be influenced by the nature of the surface. A thick layer of low density dust might serve to reflect only a small amount of energy from the surface with additional reflections taking place from a solid layer beneath the dust.

Polarization Effect

For a linearly polarized transmitting antenna, the polarization at another location may be determined for two earth stations in the Northern Hemisphere north of the sublunar geographical position (GP), the first location observing after lunar transit and the second location observing before transit, by the following equation:

$$p = \arctan\left[\cot\frac{A}{2}\left(\frac{\cos\frac{L_x-L_m}{2}}{\sin\frac{L_x+L_m}{2}}\right)\right] - \arctan\left[\cot\frac{A}{2}\left(\frac{\sin\frac{L_x-L_m}{2}}{\cos\frac{L_x+L_m}{2}}\right)\right] + \arctan\left[\cot\frac{B}{2}\left(\frac{\cos\frac{L_y-L_m}{2}}{\sin\frac{L_y+L_m}{2}}\right)\right] - \arctan\left[\cot\frac{B}{2}\left(\frac{\sin\frac{L_y-L_m}{2}}{\cos\frac{L_y+L_m}{2}}\right)\right] \quad (1)$$

where

p = polarization
L_m = sublunar latitude
L_x = latitude of the first station
A = difference in longitude between the lunar ground position and the station
L_y = latitude of the second station
B = difference in longitude between the lunar ground position and the station

If the linearly polarized receiving antenna is rotated with respect to the polarization of the incoming wave, the signal strength will be reduced by the cosine of the angle between the polarization of the antenna and that of the incoming wave. Thus a change in signal level caused by polarization effects will be encountered during a measurement period as the sublunar latitude and longitude vary as a function of time.

Faraday Rotation

The polarization of the radio wave going through the ionosphere will be

---

Paper 60-518, recommended by the AIEE Radio Communication Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Southwest District Meeting, Houston, Tex., April 4–6, 1960. Manuscript submitted December 28, 1959; made available for printing February 10, 1960.

A. W. STRAITON and C. W. TOLBERT are with The University of Texas, Austin, Tex.

This work was possible only because of the excellent co-operation of the Radio Astronomy Group at the Royal Radar Establishment in England, who provided the essential transmission as well as many valuable suggestions concerning the reception technique.

The work conducted by the Electrical Engineering Research Laboratory of The University of Texas described in this paper was sponsored by the Wright Air Development Center under Air Force Contract *33(616)-6257*.

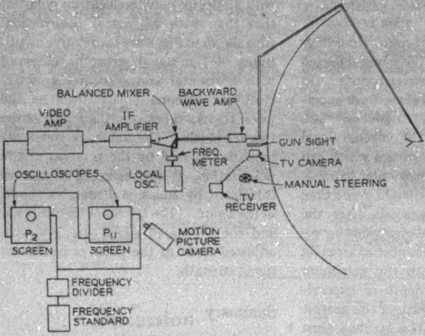


Fig. 1 (left). Receiver block diagram

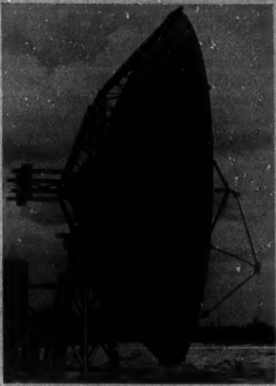


Fig. 2 (right). Receiving system

rotated through an angle α due to the movement of the charges by the wave in the earth's magnetic field. The direction of rotation will be additive for the return trip. The amount of this rotation is given for radar frequencies greater than 100 mc per second by

$$\alpha = \frac{e^3}{2\pi m^2 c^2 f^2} \int_{\ell}^{0} N(s) H(s) \cdot ds \qquad (2)$$

where e and m are the charge and mass of the electron, c is the velocity of light, and f is the radar frequency. $N(s)$ and $H(s)$ are the electron density and magnetic field, respectively, along the transmission path ℓ over which the integral is taken.

It is noted that this effect varies inversely as the square of the frequency. Extrapolating the results of vhf measurements indicates that this effect would be very small at 3,000 mc per second.

Transmitted Data

The transmitting station was located at Malvern, England, at a latitude of 52.1°N and a longitude of 2.3°W. The transmitting antenna was a 45-foot parabola which transmitted 5-microsecond pulses at a repetition rate of 250 per second. The transmitted frequency was very nearly 3,000 mc, with a power of 2 megawatts. By prearranged agreement, the repetition rate was maintained at 250 per second with a deviation of less than one part in 10^8.

Previous monostatic measurements at Malvern had shown a reflected signal strength at the receiver of 10^{-13} watts.

Receiver Data

The receiver for the studies was located at the Balcones Research Center of The University of Texas, at a latitude of 30.4° N and a longitude of 97.7° W. This is equivalent to an hour-angle separation of 6 hours and 22 minutes from Malvern and a great circle distance of 4,780 statute miles.

The receiving system, shown in block diagram in Fig. 1, was designed and assembled primarily to detect and identify the moon-reflected signal as transmitted from Malvern. The antenna and the r-f (radio-frequency) amplifier are the "heart" of the system, the other components being designed to be complementary and to display the received signal. The skin of the 28-foot antenna, a parabola formed by eight sections of Fiberglas-reinforced copper foil, was supplied by Wright Air Development Center. A framework of aluminum and steel trusses was fabricated that allowed the skin to be attached to an *F/SCR-268* radar mount. The resulting altazimuth mount was optically pointed with a telescope gunsight. Tracking of the moon was manual and was facilitated through the use of a television camera and receiver aligned with the gunsight. The parabolic reflector was horn-illuminated with a 10-db (decibel) taper and the waveguide feed from the horn terminated at the r-f amplifier. The resulting gain and beamwidth were calculated to be 46 db and 0.8 degree, respectively. A Hughes Products *PAS-2* backward-wave amplifier (bwa) was attached to the antenna mount as shown in Fig. 2, and the coaxial cables to its input were utilized to form the vertical rotary joint. The output from the bwa, operated with a gain of 20 db and a noise figure of 5 db, was waveguide-coupled to a balanced mixer and thence to an i-f amplifier with a bandwidth of 400 kc. The local oscillator, a *6BL6* klystron, was used to heterodyne to the i-f frequency of 44.5 mc, with no frequency stabilization being employed other than that inherent in the system.

The output of the second detector, after passing through a video amplifier, was applied to two cathode-ray oscilloscopes, one with a P_{11} screen for photographic purposes and the other with a P_2 screen for eye observations. Synchronizing pulses with a repetition rate of 250 ± 10^{-6} sec^{-1} were derived from a Hewlett-Packard frequency standard. Interval markers at 20 seconds were also derived for timing purposes. Provision was made for recording with a 16-mm motion picture camera, a Type A signal display, and a display similar to that used at Malvern, where the signal is applied to the z axis of an oscilloscope whose x and y coordinates are 80 seconds and 1,000 microseconds. The 1,000-microsecond sweep was synchronized with the pulse repetition rate.

The beamwidth and the pointing of the antenna were measured by means of a signal generator on top of a 270-foot tower at a distance of 2,500 feet. Observations were also made using the sun as a signal source. Prior to the measurements, the Royal Radar Establishment provided the transmitter frequency within ±0.1 mc and the periods during which the transmitter would be operated.

Measurements

The first attempt to receive the signal, made on August 22, 1959, with vertical antenna polarization, was unsuccessful. Failure was due, in part, to the antenna polarization and to a difference in indicated transmitter frequency as measured at Malvern and at Austin. The second attempt was made on the following night with horizontal polarization. Approximately 2 hours after moonrise, the signal was tuned in. The adjustable parameters of the receiver were optimized and eye observations and photographs of the intensity modulation display were made. Stratus clouds prevented optical sighting of the moon on the following night, thereby prohibiting the pointing of the antenna. The fourth night of the prescheduled measurements was canceled by mutual agreement, primarily due to the poor weather conditions at Austin but also to allow antenna repairs to be made at Malvern.

A second set of measurements was scheduled for the period between September 21 and September 24. Because of the narrow latitude permitted in adjusting the parameters of the intensity modulation scheme with the low signal-to-noise ratio experienced during the previous measurements, motion picture recordings of the Type A presentation were made during this period. No difficulty was experienced in tuning in the signal during this period. On September 21, photographic recordings and eye observations of the signal were made for approximately 5 hours with 45-degree counterclockwise antenna polarization. Low stratus clouds prevented antenna pointing on September 22. Measurements were made for approximately 5 hours with horizontal polarization on September 23. The measurements scheduled on September 24 with vertical polarization could not be made because of cloud cover.

Results

Shown in Fig. 3 are typical recordings of Type A displays of the moon-reflected signal at four different sweep speeds. On each frame of the film there were superimposed approximately 10 sweeps. The observed upper 6 db of the reflected signal has cumulative characteristics which can best be described as being similar to noise. Signal returns were distributed over a time interval of 350 microseconds after the earliest return. This corresponds to receiving signals from a number of areas on the moon's surface with the maximum signal coming from a depth of approximately 7.5 kilometers back of the nearest point on the moon. The greatest time delay corresponded to a depth of approximately 50 kilometers, which would occur at about one third of the distance from the center of the edge of the optical disk. The interval time during which the upper 6 db of a single discrete return could be observed averaged approximately one-quarter second. The mean distribution of signal amplitudes versus time are shown in Fig. 4.

The spectra which were enhanced by frequency modulating the sweep repetition rate about 20 microseconds at 2 cycles per second were primarily determined by eye observations of the P_2 screen. A more extended period of returns was observed with 45-degree counterclockwise polarization than with horizontal polarization. Also shown in Fig. 4 are the relative amplitudes of the signal strengths observed with horizontal and 45-degree counterclockwise antenna polarizations as the moon traversed between the zenith at the two terminals. The values of signal attenuation caused by the rotation of the polarization of the reflected signal as calculated by equation 1 are also shown in the figure. The curves of the observed and the calculated signal strength were adjusted to fit at the point of maximum signal strength.

Summary

The moon-reflected signal as transmitted from Malvern, England, was easily received in Austin, Texas, with a 28-foot antenna and a receiver with a noise figure of approximately 6 db. The upper 6 db of the spectra of the reflected signals were made up of a large number of pulses of varying amplitudes and positions along a time base of approximately 350 microseconds. The maximum returns occurred approximately 50 microseconds after the first return. The average polarization of the reflected signal was essentially linear and has a value which correlates reasonably well with that calculated for a plane reflecting surface. The complex nature of the reflection prevented any conclusions other than the those relating to the effective reflecting areas of the moon and to the average polarization.

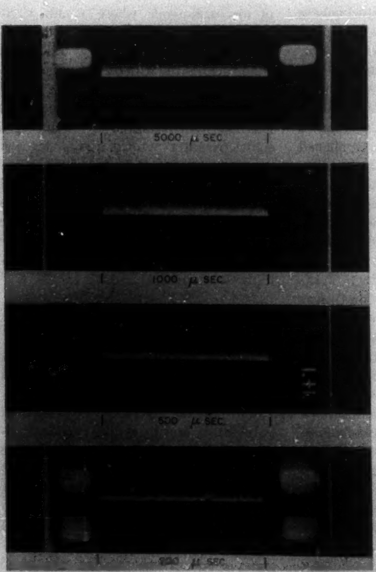


Fig. 3 (left). Type A presentation of moon-reflected signals

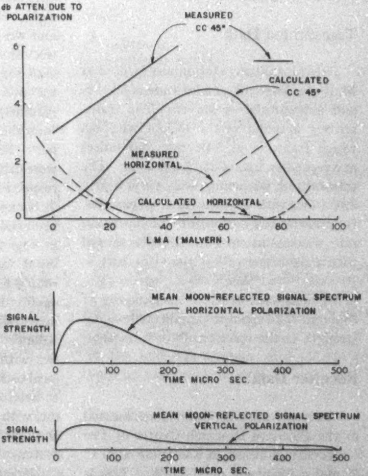


Fig. 4 (right). Spectra and polarization of reflected signals

Appendix. Signal-to-Noise Ratio

The signal-to-noise ratio for echoes from the moon is given by

$$\frac{Ps}{Pn}=\frac{P_T A_R G_T \sigma}{16\pi^2 R_T^2 R_R^2\ FKTb}$$

where

Ps = signal power at receiver
Pn = noise power
P_T = transmitter power
A_R = effective area of receiving antenna
G_T = gain of transmitting antenna
σ = scattering cross section of target
F = receiver noise
K = Boltzmann's constant
T = receiver ambient temperature
b = receiver bandwidth
R_T = distance from transmitter to moon reflecting region
R_R = distance from receiver to moon reflecting region

References

1. LUNAR RADIO ECHOES, J. H. Trexler. *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 46, 1958, p. 286.

2. THE SCATTERING OF RADIO WAVES BY THE MOON, J. V. Evans. *Proceedings*, Physical Society, London, England, vol. 70, pt. B, 1957, p. 1105.

3. RADIO ECHOES FROM THE MOON AT A WAVELENGTH OF 10 CM, B. S. Yaplee, R. A. Burton, J. K. Craig, N. C. Roman. *Proceedings*, Institute of Radio Engineers, vol. 46, 1958, p. 293.

4. RADAR MOON REFLECTION STUDIES AT 3000 MCS, V. A. Hughes. "Paris Symposium on Radio Astronomy," International Scientific Radio Union, Stanford University Press, Stanford, Calif., 1959.

5. STUDIES IN RADAR CROSS SECTION XXVIII. "The Physics of Radio Communications Via the Moon," M. L. Barash, H. Brysk, B. A. Harrison, T. B. A. Senior, K. M. Siegel, H. Weil. *Technical Report*, Engineering Research Institute, University of Michigan, Ann Arbor, Mich., Contract AF 30(602)-1725, Mar. 1958.

◆

Discussion

V. A. Hughes (Royal Radar Establishment, Malvern, Worcestershire, England): Since the first radar reflections were obtained from the moon by the United States Signal Corps and independently by Bay in Hungary in 1946, measurements have been considerably extended. The early measurements were made with comparatively low power transmitters and, in order to obtain the required signal-to-noise ratio, long pulse lengths and CW were used, producing a signal scattered back to the receiver which consisted of contributions from the whole of the lunar surface. At the meter wavelengths used, apart from a long period of fading due to Faraday rotation in the ionosphere, lunar libration produced a short period amplitude fluctuation by introducing Doppler shifts in the scattered signal from different elements on the surface. It was thought that the lunar surface would scatter isotropically, but by carrying out an analysis of the fluctuations, Evans, as pointed out in reference 2 of the paper, showed that most of the radiation was returned from a central region of the disk that had a radius only one-third of the lunar radius.

With the more recent use of radars with high peak power and microsecond pulse length it has been possible to confirm that at least from wavelengths of 1.5 meters to 10 cm this very directional type scattering exists. This has been discussed in references 1, 3, and 4 of the paper.

Measurements at Malvern at the wavelength of 10 cm using the 45-foot-diameter radio telescope with a 2-megawatt radar with 5-microsecond pulse length have shown in fact that for unit area of the lunar surface the ratio of scattered power to incident power flux, defined as the scattering cross section per unit area, is given by[1]

$$\sigma'=\sigma_0'\ exp\,\{-10\theta\}$$

where σ_0' is about 0.7, for angles of incidence θ between about 3 degrees and at least 14 degrees.

Such highly directional scattering led Senior and Siegel[2] to suggest that the echo was caused by specular reflection from three or four areas on the surface that were smooth in comparison with the wavelength; the moon has been described as quasi-smooth. The theory required that the leading edge of the echo was due to reflection from one of the regions and hence must remain nearly constant in amplitude. However the short pulse length measurements at Malvern show a leading edge which fluctuates in a random fashion; comparison with results obtained by Yaplee using a 2-microsecond pulse length shows that the mean amplitude of the leading edge is proportional to pulse length and hence to the area illuminated on the lunar surface. Moreover, a surface with irregularities much greater than a wavelength, such as the surface of the sea, also scatters in a highly directional way[3] with a law comparable to that for the moon but dependent on the state of the sea.

To decide further if the lunar scattering is from a few specular areas or from a rough surface this joint experiment was arranged between Malvern and Texas. Signals transmitted from Malvern were received both back in Malvern and also in Texas, and in the latter case were received at an angle of about $1^1/_2$ degrees to the incident angle, this being the angle subtended by the base line Malvern–Texas at the moon. If the echo were due to a few specular reflections from smooth surfaces on the moon then a correlation might be expected between the spikes in the signals received at both sites. In fact no correlation was observed; whereas the signal received at Malvern shows a steep rise at the leading edge, that received in Texas shows a gradual increase as would be expected if the scattering was from a rough surface.

The model that is suggested for the lunar surface, based on the scattering law, is one with vertical irregularities which have an rms value greater than a wavelength and a horizontal scale which is about 20 times greater. The surface would appear to be undulating and over representative annular areas of 8,000 km², corresponding to the area of the lunar surface from which signals are received with a 5-microsecond pulse, have a mean gradient of about 1 in 20. If this had been steeper, then we would have expected a more isotropic scattering, while a smaller gradient would have shown more directional reflecting properties.

This bistatic experiment has also shown the feasibility of transmitting signals for long distances over the surface of the earth, and has demonstrated the limitations that would be imposed on the use of the moon as a radio relay. Due to the type of scattering from the moon, the frequency bandwidth for a normal communication system would probably be limited to about 3 kc but a much smaller bandwidth may be necessary if good signal-to-noise ratio is required due to the "persistence" of the returned signals. Hence, to estimate the total scattering cross section of the moon for communication purposes we can integrate σ over the whole of the surface

$$\sigma=2\pi a^2\sigma_0'\int_0^{\pi/2}\sin\theta\ exp\{-10\theta\}\,.d\theta$$

$$=2\times10^{-2}\pi a^2\sigma_0$$

where a is the lunar radius, 1,738 km. For the wavelength of 10 cm, this becomes

$$\sigma=1.4\times10^{-2}\pi a^2$$

$$=1.3\times10^4\ \text{km}^2$$

However, some more recent measurements with higher performance radars and long pulse lengths at a wavelength of about 0.75 meter have shown that for larger values of θ the scattering from the lunar surface becomes isotropic with a derived scattering cross section per unit area σ_i of about 6.10^{-3}. Though there appears to be a slight wavelength dependence on the scattering, to a first approximation we shall assume that the isotropic term has the same value at the wavelength of 10 cm. The value of the scattering cross section may then have to be increased by a further factor σ_I which for the case of CW or pulse lengths greater than about 11 ms is

$$\sigma_I=2\pi a^2\sigma_i$$

$$=1.2\times10^{-2}\pi a^2$$

$$=1.1\times10^5\ \text{km}^2$$

The total scattering cross section of the moon at the wavelength of 10 cm would then be 2.4×10^5 km². The scattering cross section of the moon for pulse lengths less than 11 ms may be obtained by integrating σ and σ_i over the appropriate areas of the lunar surface.

REFERENCES

1. THE ROUGHNESS OF THE MOON AS A RADAR REFLECTOR, V. A. Hughes. *Nature*, London, England, vol. 186, 1960, p. 873.

2. RADAR REFLECTION CHARACTERISTICS OF THE MOON, T. B. A. Senior, K. M. Siegel. Paris Symposium on Radio Astronomy, International Astronomical Union and International Scientific Radio Union, Stanford University Press, Stanford, Calif., 1959, p. 29.

3. BACK SCATTERING FROM WATER AND LAND AT CENTIMETER AND MILLIMETER WAVELENGTH, C. R. Grant, B. S. Yaplee.

Proceedings, Institute of Radio Engineers, New York, N. Y., vol. 45, 1957, p. 976.

A. W. Straiton and **C. W. Tolbert.** Cross-correlations of the reflected signal as received in Austin and in Malvern were made subsequent to the writing of the paper. The correlation coefficient was found to be extremely low for the single pulse spectra. Low correlation coefficients have been reported by R. B. Dyce and R. A. Hill for diversity reception of moon reflected signals at a frequency of 106.1 mc.[1]

T. B. A. Senior and K. M. Seigel[2] find that the moon reflected signal can be explained by specular reflection from 20 or 30 major areas on the moon's surface. Those areas nearest the center of the moon producing the first of the spectrum of returns are thought to be spherical portions of the surface. Areas further back are thought to be craterlike. The individual returns constituting the reflected signal spectra per pulse of the quasi-smooth model, are each produced by a single scattering area. Rough scattering of much smaller amplitude is permitted for the total lunar hemisphere illuminated.

V. A. Hughes, as explained in his discussion, finds that the reflected signal can best be explained by a lunar model of many reflecting areas. These reflecting areas result from an undulating lunar surface. The individual pulses of the reflected signal spectrum are produced by interference between the signal reflected from these many areas.

These theoretical models are the two most commonly expounded. Either can be used to explain the low correlation of bistatic reception. The reflecting areas of the quasi-smooth model need only to be large with respect to the wavelength ($>$100 meters^2 at 3 kmcs) in order that the reflected flux per reflecting area be confined to a beam width narrow with respect to the angle subtending the receiver separation. The "undulating" model would inherently produce low correlation for bistatic reception because of the interference of the reflected signals from the many reflecting areas postulated. The gradual increase in the amplitude of the mean reflected signal spectrum received at Austin as opposed to the steep rise of the leading edge of the signal received at Malvern, along with the 3 to 6 db lower signal strength experienced at Austin relative to that received at Malvern, are more easily explained by the undulating model than by the quasi-smooth model.

Of the four parameters of the moon-reflected signal obtained from the measurements at Austin: spectrum amplitude distribution; absolute signal strength; relative amplitude versus antenna polarization; and cross correlation coefficients of the Malvern and Austin signal spectra, the absolute signal stength was subject to error of the order of the discrepancy measured and was therefore not included in the paper. The authors are however sufficiently confident of the measurement and the reference absolute level is, therefore, included.

REFERENCES

1. LUNAR ECHOES RECEIVED AND SPACED RECEIVERS AT 106.1 MC., R. B. Dyce, R. A. Hill. *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 48, May 1960, pp. 934–35.

2. A THEORY OF RADAR SCATTERING BY THE MOON, T. B. A. Senior, K. M. Siegel. *Journal of Research*, National Bureau of Standards, Washington, D. C., May 1960.

The REA-1 Transposition System

A. G. CHAPMAN
MEMBER AIEE

J. V. BUSCEMI
MEMBER AIEE

J. M. FLANIGAN
ASSOCIATE MEMBER AIEE

THE *R1* transposition system (sometimes coded *R1-5*) has been employed in many cases for transposing rural open-wire telephone lines. These have been both subscriber lines and trunk lines, voice and carrier. The *R1* system has the attractiveness of:

1. Patterns lending themselves to simplicity in engineering and construction.
2. No transposition "sections" or "*S*" poles.
3. No quantitative limits on sag or span length deviations.
4. Up to 5 pairs per crossarm and up to 4 crossarms.

The *R1* transposition system was designed primarily for use in the voice-frequency range and the available data indicate it is not too suitable for open-wire lines over which a number of multichannel carrier current systems are operated. Crosstalk considerations limit the maximum permissible carrier frequency and the number of pairs of wire suitable for carrier operation.[1,2] Unpublished data on measurements taken by Rural Electrification Administration (REA) engineers over a frequency range of 40–495 kc have since further confirmed the limitations of the *R1* transposition system for multisystem carrier application. Some representative data on the *R1* transposition system are presented later in this paper.

In looking over available transposition systems during 1954 and later, not only was the *R1* system unsuited to broader carrier application, but none were known to REA which retained much of the simplicity and economy of the *R1* system and which were designed for low-cost carrier systems having frequencies well above about 150 kc.[3–5] It was thought that open-wire carrier would become increasingly attractive economically in bringing modern telephone service to farm and other rural people as a greater variety of economical types of carrier equipment became available. What was needed was a transposition system designed to furnish a suitable transmission medium, from a noise, crosstalk, and absorption loss point of view for these new carriers. These systems have achieved their lower cost by using several times the minimum spacing of 4 kc between adjacent carrier frequencies. Through 1958, the number of carrier channels used in REA borrowers' systems were about 1,000 for trunks and 900 for subscriber lines. There is a trend toward increased use of subscriber carrier channels.

Objectives of the New System

Considering that the average REA borrower's central office has in the order of 130 lines and the average number of subscribers per borrower is around 1,200, a new transposition system did not have to provide for many crossarms and wire pairs. Nor do REA borrowers, by and large, operate long intertoll lines. Homes or central offices in rural areas may fall anywhere along a pole line route, and if such a system were to be suitable for subscriber or for trunk carrier, transposition sections with *S* poles would sharply limit its application. Borrowers in the High Plains area of the United States, and elsewhere, were finding that tandem transposition brackets had to be superseded by point brackets because the former tended to make bare-line wires subject to open-wire contacting during exceptionally windy periods. Therefore, the use of either tandem or point transposition had to be provided for.

Paper 60-43, recommended by the AIEE Wire Communication Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., January 31–February 5, 1960. Manuscript submitted November 2, 1959; made available for printing March 7, 1960.

A. G. CHAPMAN and J. M. FLANIGAN are with the Rural Electrification Administration, the former at South Orange, N. J., and the latter at Washington, D. C.; J. V. BUSCEMI is with the Anaconda Wire & Cable Company, Hastings-on-Hudson, N. Y.

The authors wish to acknowledge that they were greatly helped by reference 3 by W. C. Babcock, E. Rentrop, and C. S. Thaeler. The design and field trial of the REA-1 transposition system necessitated the co-operation of a number of REA engineers and REA borrowers' representatives. The authors are particularly indebted to REA engineer Paul G. Lambidakis, for his diligent and skillful work in making extensive calculations and field measurements.

Vol.
Pt. 1
1960

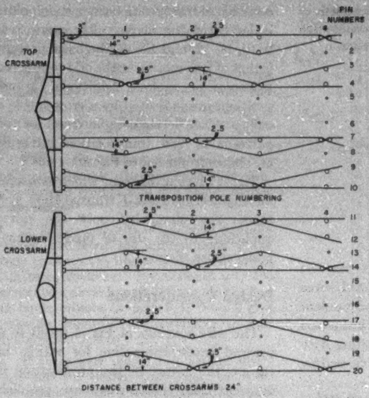


Fig. 1. Wire spacings, tandem-type brackets, type 10A crossarms

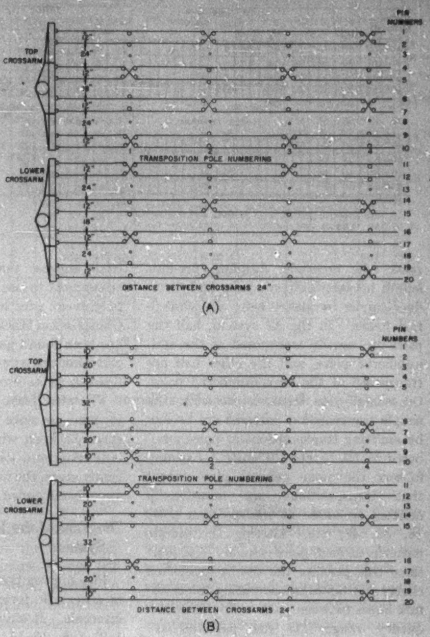


Fig. 2 (right). Wire spacings, point-type brackets
A—Type 10A crossarms
B—Type 10B crossarms

Of the entire material cost of an open-wire line, pole costs have proved to be the single biggest portion, so it was important to have a transposition system permitting average pole spacing, which would take advantage of the strength properties of bimetallic conductors possessing low carrier attenuation properties. Furthermore, a stringent tolerance on pole spacing deviations would lead to too many "set-in" poles, and therefore a greater number of average poles per mile; or tight deviation requirements would lead to midspan brackets, which may flop over or cause excessive sag deviations on longer spans. With these considerations in mind, the following objectives were established for the new transposition system:

1. A maximum of 2 crossarms and 4 pairs on each.
2. For 2 or more top crossarm pairs, a line length up to 30 miles; for second-arm pairs, about half this length.
3. The system to be "splittable," i.e., pairs could enter or leave the line (assuming proper carrier level co-ordination) at any pole.
4. Ability to use standard 10-ft (foot) *A*-type crossarms—12-in. (inch) spacing.
5. Average span lengths up to 310 ft.
6. Liberal transposition pole spacing deviations as per the rule that: the sum of all values of U^2 is not greater than $6L$, where U is the deviation, in feet, of a transposition inteval from the average length of all intervals; L is the length of the line in feet. For existing lines, the limiting value of $6L$ may be necessary to avoid excessive pole setting. For new lines a value of $3L$ or less is often readily obtained and this is highly desirable to improve crosstalk and noise.
7. Loss as a result of absorption of energy by nearby pairs not to become excessive below about 350 kc when the average transposition interval is 300 ft. For other interval lengths the top frequency is in inverse proportion to 300 ft.
8. Bare-line output-to-output far-end pair-to-pair crosstalk loss of 50 db (decibels) or more at as high a carrier frequency as possible. (This would be the minimum far-end crosstalk loss between the voice-frequency drops of like zero loss carrier channels.) It was realized that a bare-line far-end loss of 50 db or more could not be realized above about 200 kc and that in the range of about 200–350 kc, it would be necessary to depend on the crosstalk advantages inherent in the design of various carrier systems. These advantages are attributable to compandors, staggered carrier frequencies, frequency inversion, FM (frequency-modulated) capture, etc.
9. Bare-line input-to-output near-end pair-to-pair crosstalk loss of not less than about 40 db. The crosstalk currents involved in such a loss are reflected at points of impedance mismatch and appear as far-end crosstalk currents. As in item 8, crosstalk advantages are needed to meet the near-end loss requirements at the higher frequencies.

Thus, the *REA-1* transposition system was developed with the intent of retaining much of the simplicity and economy of the *R1* system. At the same time steps were taken to provide marked improvement in carrier frequency crosstalk losses and noise and consequent simplification in engineering a number of carrier systems on the same open-wire line. Because low-cost carrier systems have carrier frequencies up to about 350 kc,[6–9] the frequency range up to 500 kc was considered in design and in field measurement of crosstalk losses between pairs and insertion losses between the two ends of each pair.

The physical nature and engineering of the *REA-1* transposition system are discussed in detail in the REA Telephone Engineering and Construction Manual—463.[10] A brief discussion of the transposition system performance and some of the data obtained in a field trial will be given in this paper. For background information, general design considerations involved in the development of the *REA-1* system will be touched on.

Special Features of REA-1 Transposition System

Unlike the *R1* system, the pole pairs are omitted. This was done to im-

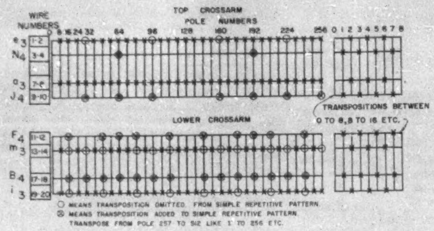

WIRE NUMBERS
TOP CROSSARM
POLE NUMBERS
0 8 16 24 32 64 96 128 160 192 224 256 0 1 2 3 4 5 6 7 8
e3 1-2
N4 3-4
a3 7-8
J4 9-10
TRANSPOSITIONS BETWEEN 0 TO 8, 8 TO 16 ETC.
LOWER CROSSARM
F4 11-12
m3 13-14
B4 17-18
i3 19-20
○ MEANS TRANSPOSITION OMITTED FROM SIMPLE REPETITIVE PATTERN.
⊗ MEANS TRANSPOSITION ADDED TO SIMPLE REPETITIVE PATTERN
TRANSPOSE FROM POLE 257 TO 512 LIKE 1 TO 256 ETC.

Fig. 3. Schematic of REA-1 transposition system

prove the crosstalk coupling per-unit length between different pairs and to reduce carrier frequency noise and absorption losses. In the *R1* system, half the pairs are usually transposed at the odd-numbered poles, and the other half are transposed at the even numbered poles. Occasionally, to improve crosstalk, the simple transposition patterns are broken by omitting transpositions at poles numbered 25, 50, 75, etc. The *REA-1* system follows the same general plan except that transpositions are occasionally omitted or added at transposition poles 16, 32, 48, etc. Altering the simple patterns at such poles is highly expedient since this permits the use of published charts and tables in computing the crosstalk losses between pairs over a wide frequency range.[3] If the patterns are altered at poles 25, 50, 75, etc., the computations are simple at voice frequencies but very involved at carrier frequencies.

Unlike the *R1* system, the rules for the *REA-1* system require control of the differences in sag of the two wires of a pair (±2 in., initially) and control of the differences in transposition pole spacing. As a result of the required transposition pole spacing control, every pole may not be a transposition pole (i.e., a pole at which some pair is transposed).

Transposition Brackets

Fig. 1 indicates schematically the wire configuration when tandem transposition brackets are used with the *REA-1* system. The figure shows the first four transposition intervals. The wires of a pair taper from 14 to 2.5 in. in each transposition interval or vice versa (3 in. at the dead-end pole). If there is an intermediate pole in a transposition interval, its pinholes are specially bored to place the insulators in position to maintain about the normal taper in a transposition interval.

Tandem transposition brackets are used in moderately windy areas for economy. In exceptionally windy areas 12-in. point-type transpositions may be used as indicated in Fig. 2(A). Limited REA field trials suggest that someday it may be economical to use plastic insulated wires to prevent wire hits and to use tandem transposition brackets. Plastic insulation has been found useful in preventing wire corrosion and wire-to-wire leaks caused by Spanish moss, etc.

The wire spacings of Fig. 1 are suitable for joint use since there is ample climbing space between wires 4 and 7, etc. For bare wires and joint use in exceptionally windy areas, the wire spacings of Fig. 2(B) are recommended.

Diagram of the REA-1 Transposition System

Fig. 3 gives the transposition patterns in schematic form for 256 transposition intervals. It may be seen that for pairs 1–2, 7–8, 13–14, and 19–20 the basic pattern is that of transposing at the even-numbered transposition poles. The omission of transpositions at certain points is indicated by circles. For pairs 3–4, 9–10, 11–12, and 17–18, the basic pattern is that of transposing at the odd-numbered transposition poles. Additional transpositions are indicated by a cross inside a circle. The "letter designations" for the patterns in 256 transposition intervals are shown to the left of Fig. 3.[3]

Fig. 4 gives a diagram which is useful for engineering *REA-1* transposition systems. For simplicity only the notes applicable to the use of tandem transposition are given.

Design Considerations

The wire spacings of Fig. 1 result in a tapered transmission line for which, in theory, the nature of the propagation is different than if the wires are parallel. However, the slightly tapered transmission line of Fig. 1 behaves about like one in which the wires are parallel and about 8 in. apart. For 300-ft transposition pole intervals and about 405 kc, the phase change per transposition interval is about 45 degrees and there is a large attenuation-frequency "absorption peak."

Referring to Fig. 3, in 256 transposition intervals there are two basic patterns a_3 (transpositions at even-numbered poles), and P_4 (transpositions at odd-numbered poles). If there were no circles or circle crosses, half the pairs would be transposed a_3 and the other half P_4.

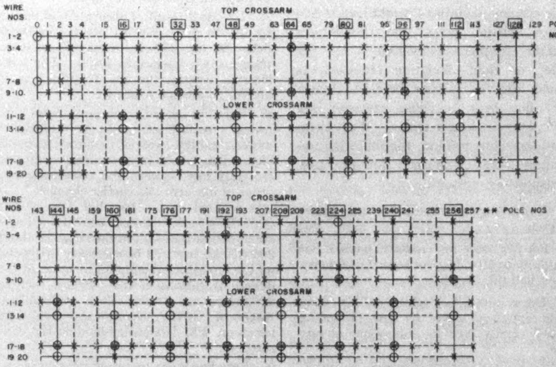

WIRE NOS
TOP CROSSARM
0 1 2 3 4 15 16 17 31 32 33 47 48 49 63 64 65 79 80 81 95 96 97 111 112 113 127 128 129 POLE NOS
1-2
3-4
7-8
9-10
LOWER CROSSARM
11-12
13-14
17-18
19-20
WIRE NOS
TOP CROSSARM
143 144 145 159 160 161 175 176 177 191 192 193 207 208 209 223 224 225 239 240 241 255 256 257 ** POLE NOS
1-2
3-4
7-8
9-10
LOWER CROSSARM
11-12
13-14
17-18
19-20

Fig. 4. REA-1 transposition system, tandem brackets

○ Transposition omitted. Use double groove insulator (type TW) to pinch wires to 2.5-in. spacing, except for dead-end poles where clevises are used to pinch to 3 in.

⊗ Point-type transposition

✕ Tandem transposition

** Transpose from pole 257 to 512 like 1 to 256 etc. At transposition poles with boxed numbers, 16, 32, etc., special transposition patterns are used. For last pole substitute ○ for ✕ and omit ⊗

To implement the objectives for the new system, the following design considerations were kept in mind:

1. Pairs transposed alike for a considerable distance may have objectionable near-end crosstalk at voice frequencies and objectionable far-end crosstalk at the high carrier frequencies.

2. Pairs transposed to patterns a_4 and P_4 for a considerable distance have objectionable far-end crosstalk at the higher carrier frequencies.[19]

3. If 256 transposition intervals actually existed, the fewer the interruptions of the basic patterns the higher (better) the crosstalk losses would be. For example, with 256 intervals, a circle in pair 7–8 at pole 128 and no circle crosses in pair 9–10 would give desirable crosstalk characteristics to pair combination 7–8 to 9–10.

4. If there were only a few pattern interruptions, the design could not be split. For example, if the line length has 127 transposition intervals, an interruption in basic pattern at only transposition pole 128 does not break up the basic patterns in 127 intervals.

5. Too many circle crosses are undesirable since, when tandem transpositions are used as much as possible, the circle crosses represent relatively expensive point-type transpositions. From Fig. 3 and about 310-ft transposition pole spacing it may be calculated that the transpositions per pair mile are:

Crossarm	Tandem Transpositions	Point Transpositions
Top	8.5	0.15
Second	8.2	0.40

FIG. 6 CASES	WIRE NO.	1-2	3-4	7-8	9-10	11-12	13-14	17-18	19-20
A	1-2	93	74	35	21	66	58	27	17
B		44	30	30	21	31	19	18	19
C		20	20	10	7	16	14	8	7
A	3-4		207	77	35	60	155	29	28
B			107	43	30	26	78	45	22
C			37	17	10	14	41	14	9
A	7-8			207	74	28	29	155	60
B				107	30	22	45	78	26
C				37	20	9	14	41	14
A	9-10				93	17	27	58	66
B					44	19	18	19	31
C					20	7	8	14	16
A	11-12					107	79	36	20
B						45	35	28	21
C						26	21	11	8
A	13-14						233	79	36
B							108	57	28
C							47	17	11
A	17-18							233	79
B								108	35
C								47	21
A	19-20								107
B									45
C									26

(A)

FIG. 6 CASES	WIRE NO.	3-4	7-8	9-10	11-12	13-14	17-18	19-20
A	1-2	974	47	27	653	92	34	23
B		380	76	23	667	105	51	18
C		218	28	10	312	46	18	9
A	3-4		133	47	31	549	66	32
B			589	76	111	622	102	46
C			140	28	44	289	61	18
A	7-8			974	32	66	549	31
B				380	46	102	622	111
C				218	18	61	289	44
A	9-10				23	34	40	653
B					18	51	105	669
C					9	18	46	312
A	11-12					979	45	28
B						410	77	23
C						224	26	10
A	13-14						130	45
B							603	77
C							133	26
A	17-18							979
B								410
C								224

(B)

Fig. 6. (A) Far-end and (B) near-end crosstalk coefficients in crosstalk units/mile/kc

Case A—Measured[3]
Case B—Computed per REA
Case C—Computed[3]

For comparison, a long (7.5-mile) 0–1 transposition section has about 12.4 point transposition per carrier pair per mile. The difficulty in choosing the circles and circle crosses depends on the number of pair combinations. Because it was felt that the top crossarm would usually be ample to give the desired number of voice and carrier channels it was decided first to design the top arm as simply and economically as possible and let the second arm be more complicated. The second arm is used for voice channels and for a relatively few carrier channels for short distances.

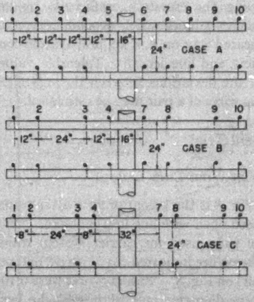


Fig. 5. Configuration of open-wire pairs on two 10-ft crossarms, wire diameter 0.104 in.

Case A—Eight 12-in. pairs and two 18-in. pole pairs
Case B—Eight 12-12-in. pairs
Case C—Eight 8-in. pairs
All wires tied out

Crosstalk Coefficients

As explained in reference 3 near-end or far-end crosstalk current ratios (times 10^6) between open-wire pairs are calculated by multiplying a "crosstalk coefficient" by an "unbalance factor." The crosstalk coefficients are expressed in crosstalk units per mile per kilocycle. The crosstalk coefficients are preferably measured but may be calculated with reasonable accuracy.[3]

For parallel wires those coefficients can be computed from the geometry of the wire configuration.

Fig. 5 indicates three wire configurations for two 10-ft crossarms 24 in. apart. Line wires of 0.104-in. diameter were assumed in computing coefficients.

(A)—Four 12-in. pairs and one 18-in. pole pair per crossarm.

(B)—Four 12-in. pairs per crossarm.

(C)—Four 8-in. pairs per crossarm.

The coefficients for case *A* were measured.[3] The coefficients for case *C* were computed.[3] The coefficients for case *B* were computed by REA engineers by the method of reference 3. As a check an alternative method of estimating coefficients was also used. This involved calculations of mutual inductance between pairs, direct capacitance unbalances between pairs and carrier frequency characteristic impedances of pairs. The direct capacitance unbalances were estimated from published measured direct capacitances for 4 to 30 parallel telephone wires.[10]

Case *A* indicates the normal wire arrangement with a type *10A* crossarm. Case *B* indicates the wire configuration used with *REA-1* transpositions in especially windy areas; see Fig. 2(A). Fig. 1 shows the wire configuration used with *REA-1* transpositions in moderately windy areas where tandem transpositions are used. The mean wire spacings, in inches, of Fig. 1, in any span are about:

8 22 8 38 8 22 8

The wire spacings of case *C* of Fig. 6 are, in inches:

8 24 8 34 8 24 8

It was hoped that the crosstalk coefficients for the wire configuration of Fig. 1 would resemble those for the wire configuration of case *C* of Fig. 5 because of the rough similarity in the previous two sets of wire spacings. The far-end and near-end crosstalk coefficients for cases *A*, *B*, and *C* of Fig. 5 are given on Figs. 6(A) and (B).

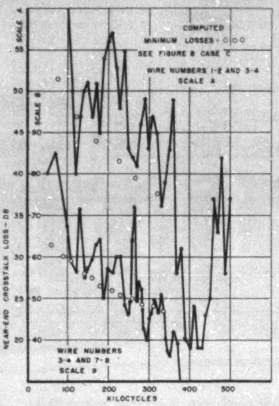


Fig. 7 (left). REA-1 transpositions. Near-end input-to-output crosstalk losses measured and computed

Computed losses due to transposition patterns: 0.104-in. 40% copperweld, 70 F

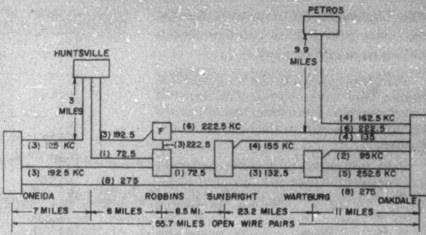


Fig. 8 (above). Open-wire carrier systems in Tennessee

Channels per pair = average 5, maximum 8
Greatest length of channel = 55.7 miles
Wires are 0.104 in. 40% copperweld
(8) 275 = 8 channels, top frequency of 275 kc
F = pole-mounted carrier-separation filter.
Open-wire carrier pairs on top arm except two are on second arm Wartburg-Oakdale

It must be seen from Fig. 6(A) that the case *B* configuration usually results in a considerable reduction in far-end coefficients from those for case *A*. Open-wire carrier channels use different frequencies for opposite directional transmission and far-end crosstalk is controlling. The lower far-end coefficients shown for case *B* than for case *A* were the justification for omitting the pole pairs. On Fig. 6(A), case *C* coefficients are considerably lower than case *B* coefficients and as noted above, it was hoped that the case *C* coefficients would apply for the tapered wire configuration of Fig. 1.

However, crosstalk data from a field trial of a two crossarm line having 128 accurately spaced 300-ft transposition intervals indicated that the far-end crosstalk coefficients for the tapered pairs lay between the case *B* and *C* values of Fig. 6(A) and usually nearer the case *B* values. This is not surprising since, strictly speaking, the crosstalk coefficients for tapered pairs vary from point-to-point in a span.

Fig. 6(A) shows the pair-to-self far-end crosstalk coefficients (1–2 to 1–2, etc.). Such coefficients control the magnitude of absorption peaks and the coupling between output and input of amplifiers via interaction crosstalk.[3] The pair-to-self coefficients are improved by a factor of about 0.5 by the use of 4 rather than 5 pairs per crossarm; case *B* versus case *A*.

Fig. 6(B) shows near-end crosstalk coefficients for cases *A*, *B*, and *C*. While the case *B* coefficients are lower than the case *A* coefficients for the larger values such as those for 1–2 to 3–4, case *B* has higher coefficients for such pair combinations as 3–4 to 7–8, 1–2 to 11–12, etc. It was hoped that the coefficients of case *C* rather than those of case *B* would apply to the tapering spans of Fig. 1. Fortunately the 128 transposition interval field trial indicated that the larger near-end coefficients were nearer those of case *C* than those of case *B*. This is indicated on Fig. 7 by a comparison of measured near-end crosstalk losses and the minimum losses computed as a result of the transposition patterns.

Unbalance Factors Due to Transposition Patterns

As noted in the foregoing, the total crosstalk in a given length is calculated, in crosstalk units (CU), by multiplying a crosstalk coefficient by an unbalance factor. The crosstalk loss in decibels = $120-20\log_{10}$ (CU).

Charts and tables are given in reference 3 for computing "type unbalance factors," that is, the unbalance factors depending on the transposition patterns. These are tabulated or plotted as kilocycle miles versus *BD* where *B* is the phase change in degrees per mile for some particular frequency in kilocycles. The length in miles for some particular transposition pattern is indicated by *D*. If *V* is the velocity in miles per second at some particular frequency in kilocycles,

$$BD = 2\,(\text{kc})(D)\,\frac{180{,}000}{V} = 2\,(\text{kc})\,D$$

approximately

The type unbalance factors are given in reference 3 for symmetrical transposition patterns called "fundamental transposition types" which may have "extra transpositions" in each 32nd of a length *D*. Type unbalance factors rise and fall with *BD* but the highest (worst) values increase rapidly with *BD*.

Unsymmetrical patterns such as those used with the *R1-5* transposition system require special calculations. Because the *REA-1* transposition system is a "split-table" design, a length of line actually used may not contain a symmetrical transposition pattern but only a part of such a pattern. Care was taken in choosing the patterns to avoid excessive crosstalk currents due to partial patterns.

Unbalance Factors Due to Wire and Pole Spacing Irregularities

Even if the unbalance factors due to transposition patterns are made very small there still may be serious crosstalk due to irregularities in wire spacing and in transposition pole spacing. As discussed in reference 3 if there are *n* transposition intervals in a transposed line and if the rms deviation from the average interval is *r* feet, it is possible to estimate the rms crosstalk due to these transposition interval deviations. This is done by multiplying the crosstalk coefficients by unbalance factors which depend in a complicated way on *n*, *r*, kc, propagation constant, and line length. Tolerable values for the deviations between the sags of the two wires of a pair may be estimated.

Field Trial

OPEN-WIRE LINE LAYOUT

Fig. 8 is the layout of an *REA-1* transposed open-wire line in Tennessee. Only the pairs used for carrier system transmission are indicated (by a single line per pair) on Fig. 8. The line wires were 0.104 in. in diameter and were 40% d-c conductivity copperweld. In all parts of the main line and its branches, there existed at least the four pairs on the top arm. The second crossarm with one to four pairs existed in short stretches. Pairs not presently used for carrier frequency trunks as shown in Fig. 8 are used for voice-frequency loops or trunks. Between Oakdale and Wartburg (11 miles) there were four carrier pairs on the top arm and two carrier pairs at the ends of the second arm. One or two other pairs existed on the second arm in short stretches.

Fig. 8 does not show the short 22-gauge nonloaded entrance cable which brought the carrier pairs into the offices. These cables were usually not longer than 0.3 kilofoot. Exceptions were a cable of 2.3 kilofeet at Wartburg and one of 3.0 kilofeet at Oneida. Pair 9–10 carries 8 carrier channels with a top carrier frequency of 275 kc from Oakdale to Oneida, a distance of about 58 miles. To avoid a loss of about 10 db in the Oneida cable at a possible channel 10 carrier frequency of 335 kc, pair 9–10 was bypassed around the cable over an insulated open-wire pair. This pair was transposed at every other pole. Each span tapered from 4-in. wires spacing to one-inch spacing. The plastic insulated wires were placed in special fittings which were mounted on small angle irons which were bolted to the cable poles a little above the cable. The measured loss of pair 9–10 at 335 kc between the Oakdale and Oneida offices was 34 db. This loss is low enough to ensure an excellent speech to noise ratio.

As may be calculated from data on Fig. 8, there are about 1,183 channel-miles used over 238 pair-miles or an average of about 5 channels per pair. These data are as of June 15, 1959. At that time the highest carrier frequency for like channels on different pairs was 222.5 kc. This was for pairs 1–2 and 7–8 in the 11-mile stretch from Oakdale to Wartburg. In June 1959, measurements of noise and crosstalk volume were made at the carrier line side of all carrier receiving terminals indicated on Fig. 8, with the test pair de-energized. These volumes were measured in decibels relative to one milliwatt (dbm) in 6-kc bands centered at the various carrier frequencies. The measured volumes were corrected for receiving gain or loss and indicated that noise or crosstalk volumes arising in the outside plant were low enough that modern telephone system objectives would be met at the voice drop terminals. This was confirmed by listening tests at the voice-frequency carrier channel drops. All channels were found to have very little noise or crosstalk volume.

Although all parts of the open-wire line indicated on Fig. 8 had not been completed in Sept. 1957 because of the right-of-way difficulties, it was decided to make carrier frequency insertion loss and crosstalk loss measurements on completed sections of line at that time. As this was the first installation of the *REA-1* transposition system, it was thought important to get an experimental check on the theoretical performance of the REA-1 system as soon as possible.

It was hoped to measure insertion and crosstalk losses for a distance of 42 miles between Oakdale and Huntsville and for various fractions of this distance. To avoid delay however the following three lengths were measured (TP indicating transposition pole),

From	To	Approximate Length, Miles	Number of Pairs Tested
Oakdale	*TP 128*	7.3	6
Robbins	Huntsville	9.2	4
Oakdale	Petros	22.0	4

In order to obtain the maximum possible number of pairs for test in these lengths, it was necessary to make low capacitance crossconnections at the Petros junction pole (TP201) and at the Huntsville junction pole (TP512 plus 194).

In the branch line from Petros junction pole to Petros, the transposition poles were numbered (and transposed) from 1 to 182. After the tests it was concluded that somewhat higher crosstalk losses would have been obtained for some pair combinations by numbering (and transposing) the transposition poles continuously from Oakdale to Petros. This would have numbered the transposition poles on the branch line from TP201 to TP382. This numbering method for branch lines was later specified in REA Manual 463. Likewise, while the Robbins to Huntsville junction pole was numbered with the appropriate main-line numbers TP512 plus 84 to TP512 plus 194, the branch line to Huntsville was numbered TP zero to TP49 instead of TP512 plus 194 to TP512 plus 243.

The average length of transposition interval (usually one span) and the deviations in these intervals for the three test lengths are as follows:

Test Section	*L*, Ft	Average Transposition Interval, Ft	Sum of U^2, Ft
Oakdale–TP128	38,534	300	0.88 *L*
Robbins–Huntsville	48,579	306	5.7 *L*
Oakdale–Petros	116,160	300 (about)	1.0 *L* (about)

In the Oakdale-TP128 section there were only 7 transposition intervals with *U* deviations which varied appreciably from the average value of 300 ft. These *U* values varied from 22 feet to 118 ft. In the Robbins-Huntsville section, because

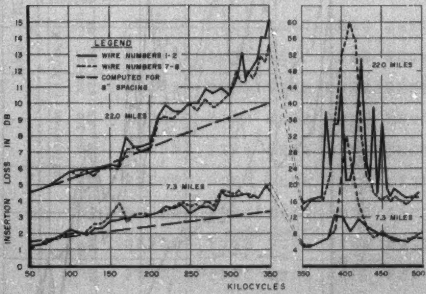


Fig. 9. REA-1 transpositions, insertion loss between 600-ohm resistors, 70 F

Tapered 300-ft spans, 0.104-in. 40% conductivity copperweld wires

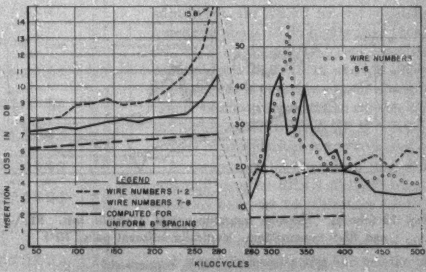


Fig. 10. Insertion loss between 600-ohm resistors in 14.1-mile length, 0.080-in. 30% copperweld

RI-5 transpositions. Weather fair, 70 F

Average span length 367 ft

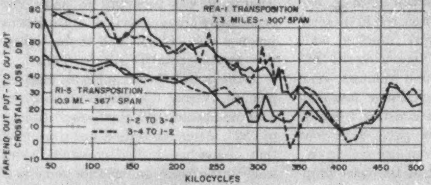


Fig. 11. Far-end output-to-output crosstalk losses. REA-1 versus RI-5 transpositions. Adjacent pairs 1–2 and 3–4, 70 F

REA-1—104-in. 40% copperweld
RI-5-0.080-in. 30% copperweld

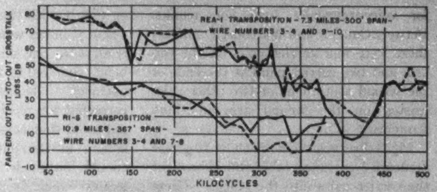


Fig. 12. Far end output to output crosstalk losses

REA-1 used 0.104-in. 40% copperweld
RI-5 used 0.080-in. 30% copperweld wires
Alternate pairs on same crossarm; 70 F

of the rugged terrain, there were 35 U values of appreciable amount. These values varied from 11–200 ft. There were 5 transposition intervals over 400 ft long (410–464 ft). These were used to span ravines and streams. The shortest transposition interval was 100 ft.

Of the three test sections, the Oakdale TP128 section was the most useful for comparison of theory with measurements because the crosstalk losses for 128 uniformly spaced transposition intervals may be computed in a reasonable time and since all combinations of 6 pairs could be measured.

INSERTION LOSSES BETWEEN 600-OHM RESISTORS

Insertion losses (in decibels) were measured for all the pairs in all three test sections. Representative insertion loss frequency curves are shown on Fig. 9 for 7.3 miles and for 22 miles. The computed loss for a single uniformly spaced 8-in. pair is shown for comparison (up to 350 kc).

Inside pair 7–8 is transposed at all even-numbered transposition poles. As predicted by theory,[3] there is a large sharp absorption loss peak at around 400–410 kc. The engineering rule is that

$$\text{kc} = \frac{240{,}000}{2d}$$

where $2d$ is the distance between transpositions in feet (in this case about 600 ft)

At the peak frequency, the absorption loss varies rapidly with length as may be seen from Fig. 9. At a somewhat lower frequency, 350 kc, the increase in loss per mile with length is not rapid. It may be seen from Fig. 9 that the loss is increasingly greater than the computed value from about 150 kc upward. This may be explained on the basis of an extra out-of phase arriving current which may be computed from pair-to-self far-end crosstalk theory.[3]

Outside pair 1–2 is transposed like inside pair 7–8 except that transpositions are omitted at transposition poles 32, 96, 160, 192, etc. As will be seen from Fig. 9, these omitted transpositions level the very high peak at 400–410 kc but they splash the debris over the kilocycle landscape and do not help in making frequencies usable above about 350 kc.

Since the top usable frequency is in inverse proportion to the transposition pole spacing, shortening the spacing increases the top frequency and vice versa. This is illustrated by Fig. 10 which shows insertion loss-frequency curves for pairs 1–2 and 7–8 of an *R1-5* transposed line and for a line length of 14.1 miles. Here the critical frequency is 320–330 kc instead of 400–410 kc because the usual spacing between transpositions in any one pair is about 734 ft instead of 600 ft. With pair 5–6 an absorption peak of 55 db at 330 kc was measured for the 14.1-mile length. This pair is transposed at each odd-numbered transposition pole. This tends to build up a large single peak which is probably somewhat reduced because the transposition interval deviations are quite large (summed $U^2 = 9L$). The other pairs

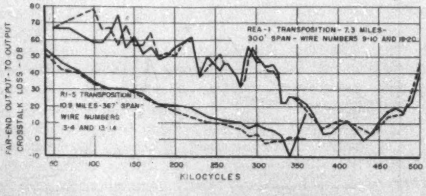
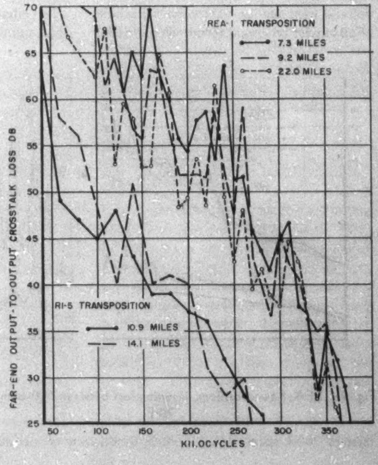


Fig. 13 (above). Far-end output-to-output crosstalk loss. REA-1 and RI-5 transpositions. Vertically adjacent pairs, 70 F

REA-1—0.104-in. 40% copperweld
RI-5—0.080-in. 30% copperweld

Fig. 14 (right). Far-end output-to-output crosstalk losses

Plotted points are averages of losses for pair 1–2 to pair 3–4 and 3–4 to 1–2, 70 F



Vol.
Pt. 1
196

have transpositions omitted at some of the transposition poles numbered 25, 50, 75, etc. These omitted transpositions tend to splash the big peak about but the usable top frequency appears to be 290–300 kc.

CROSSTALK LOSSES

In each of the test sections, all combinations of the through pairs were measured for input-to-output near-end crosstalk loss in db and output-to-output far-end crosstalk losses. Fig. 7 showed measured and computed near-end crosstalk losses (Oakdale to TP128) for combination 1–2 and 3–4 and combination 3–4 and 7–8.

Figs. 11, 12, and 13 show representative measured far-end crosstalk losses for horizontally adjacent pairs, horizontally alternate pairs, and vertically adjacent pairs. Losses when sending on the lower-numbered pair (of a pair combination) are indicated by solid lines and losses when sending on the higher numbered pair are indicated by dashed lines. The solid and dashed lines differ somewhat and this is theoretically a result of irregularities in construction.

On Figs. 11, 12, and 13, a comparison is made between *REA-1* transposed pairs and *R1-5* transposed pairs. The comparison is not an exact one since *REA-1* pairs were 7.3 miles long and had 0.104-inch 40% conductivity copperweld conductors while the *R1-5* pairs were 10.9 miles long and had 0.080-inch 30% copperweld wires. The average transposition pole spacing was 300 ft for the *REA-1* pairs and 367 ft for the *R1-5* pairs.

If the *R1-5* transposition pole spacing were reduced to 300 ft and the line length reduced from 10.9 to (300/367) 10.9, or 8.9 miles, the *R1-5* transposed pairs would be expected to have 5 to 10 db higher far-end crosstalk losses over the frequency range of 50–350 kv. Even with such an adjustment, the *REA-1* far-end crosstalk losses run about 15 db higher than those for *R1-5* transpositions. This may be deduced from Figs. 11, 12, and 13. This improvement is probably a result of selection of transposition patterns suitable for the higher carrier frequencies[8] and better control of irregularities in construction.

A comparison of measured near-end crosstalk losses for *REA-1* and *R1-5* transpositions indicates an advantage for the former of about 15–20 db.

Fig. 14 shows measured values of far-end crosstalk loss for the line lengths of 7.3, 9.2, and 22.0 miles. The pair combination is 1–2 and 3–4. At each frequency there is given the average of two values of crosstalk loss, one for sending on 1–2 and the other for sending on 3–4. As

Table I. REA-1 Transposition System, Minimum Far-End and Near-End Crosstalk Losses in Decibels

Pair Combinations	128 Intervals—7.3 Miles, Kc											
	0–100		100–150		150–200		200–250		250–300		300–350	
	FE	NE	FE	NE	FE	NE	FE	NE	FE	NE	FE	NE
1–3*	69	61	62	40	54	45	49	43	42	40	27	36
1–7*	66	54	65	52	49	51	54	51	40	45	36	38
1–9*	73	80	63	70	53	65	55	56	54	52	31	49
3–7*	56	68	66	55	48	50	52	46	45	40	38	36
3–9*	74	59	50	56	52	57	56	53	43	47	32	40
7–9*	72	63	58	45	53	45	48	47	40	43	33	33
1–11*	65	46	50	35	53	42	43	39	34	34	19	25
1–13	74	41	59	41	55	41	45	41	41	41	25	41
1–17	80	64	76	58	70	56	57	52	51	51	33	49
1–19*	73	68	64	53	51	59	47	50	45	48	25	38
3–11*	61	58	54	47	49	49	43	47	38	40	23	33
3–13	60	41	54	31	50	31	35	26	31	26	6	22
3–17	75	39	75	39	70	39	53	39	47	39	19	39
3–19*	60	75	62	67	64	59	53	71	50	47	36	40
7–11*	73	71	68	63	56	63	58	57	47	51	31	37
7–13	72	39	60	39	55	39	41	39	38	39	16	39
7–17	72	41	55	31	50	31	36	26	30	26	12	22
7–19*	65	61	60	50	52	50	43	48	42	40	34	37
9–11*	78	73	58	62	66	58	60	56	51	55	35	40
9–13	75	64	75	58	70	56	57	52	51	51	33	49
9–17	75	41	75	41	75	41	66	41	56	41	34	41
9–19*	58	44	55	35	48	38	37	37	31	31	21	25
11–13	75	47	56	42	53	39	37	37	34	35	8	33
11–17	75	48	70	48	68	48	50	48	45	48	16	48
11–19*	68	73	66	67	62	59	51	52	47	53	31	36
13–17	70	48	48	44	50	41	50	39	29	36	2	35
13–19	65	48	52	48	49	48	38	48	35	48	13	48
17–19	65	47	60	42	58	39	55	37	39	34	15	33

Pair Combinations	256 Intervals—14.5 Miles, Kc											
	0–100		100–150		150–200		200–250		250–300		300–350	
	FE	NE	FE	NE	FE	NE	FE	NE	FE	NE	FE	NE
1–3*	69	46	62	42	61	39	51	36	41	34	28	33
1–7*	66	47	65	47	57	47	50	47	42	45	31	40
1–9*	73	75	63	65	61	62	51	55	42	50	30	48
3–7*	60	47	65	43	55	41	52	37	44	37	30	35
3–9*	60	48	55	47	55	47	55	46	44	46	30	35
7–9*	60	46	50	42	50	39	50	36	46	34	34	33
1–11*	65	39	50	30	52	30	40	25	34	25	9	21
1–13	65	39	50	39	46	39	36	39	32	39	17	39
1–17	76	60	75	57	68	54	56	51	50	50	30	47
1–19*	70	52	60	52	50	52	46	52	42	50	28	40
3–11*	60	38	50	38	50	38	50	38	43	38	19	30
3–13	60	39	53	31	48	31	33	26	29	26	8	22
3–17	70	37	70	37	64	37	44	37	39	37	9	37
3–19*	70	63	65	57	64	55	53	51	47	45	23	40
7–11*	70	63	68	57	55	55	56	52	47	50	23	40
7–13	70	37	53	37	50	37	40	37	36	37	20	37
7–17	70	39	55	31	54	30	37	26	36	25	10	22
7–19*	65	38	60	38	55	38	48	38	42	38	31	35
9–11*	70	52	60	52	65	52	50	52	40	52	24	40
9–13	70	63	65	57	62	54	51	51	44	50	23	47
9–17	70	39	70	39	70	39	57	39	48	39	25	39
9–19*	60	39	55	30	48	30	38	25	30	25	18	21
11–13	70	46	55	41	49	38	35	36	30	34	10	33
11–17	70	47	70	47	65	47	41	47	37	47	6	45
11–19*	65	70	65	65	56	55	45	50	37	50	22	35
13–17	50	48	47	44	45	41	29	38	26	35	1	35
13–19	60	47	49	47	46	47	36	47	32	47	17	45
17–19	60	46	55	41	57	39	40	36	38	34	13	33

Pair Combinations	512 Intervals—29.1 Miles, Kc											
	0–100		100–150		150–200		200–250		250–300		300–350	
	FE	NE	FE	NE	FE	NE	FE	NE	FE	NE	FE	NE
1–3*	68	45	61	41	60	38	50	35	40	33	27	32
1–7*	65	46	64	46	56	46	49	46	41	44	30	40
1–9*	72	74	62	64	60	61	50	54	41	49	29	47
3–7*	59	46	60	42	54	40	51	36	43	36	29	34
3–9*	59	47	54	46	54	46	54	45	43	45	29	34
7–9*	59	45	49	41	49	38	49	35	45	33	33	32

First figures give output-to-output far-end losses; second figures give input-to-output near-end losses. 300-foot transposition intervals. A pair is indicated by its lowest numbered wire.

* Measured in 7.3-mile stretch. Other crosstalk losses are computed using case B coefficients of Figs. 6(A) and (B) and assuming sum of $U^2 = 6L$ in feet.

shown by Fig. 14, the effect of line length in decreasing the crosstalk loss is not marked even though the transposition patterns for the 9.2- and 22.0-mile lengths were not completely symmetrical ones. Also, as previously discussed, the transposition pole spacing deviations were much greater in the 9.2-mile length than in the other two lengths.

Thus, theory and experiment indicate that with *REA-1* transpositions and for combinations of pairs on the top crossarm, the crosstalk losses do not decrease with length to any marked extent. However, the rate of decrease of crosstalk losses with increase in transposition pole spacing is quite rapid. Contrariwise, the rate of increase of crosstalk losses with decrease of pole spacing is quite rapid. This effect may be deduced from Figs. 7, 11, 12, 13, and 14, by multiplying the frequency scales by $300/d$ where d is the transposition pole spacing in feet.

Crosstalk Loss Data for Engineering Purposes

For engineering purposes Table I has been prepared. This gives crosstalk loss data for pair combinations on the top crossarm and for 7.3-, 14.5-, and 29.1-mile lengths. The crosstalk losses were measured for the 7.3-mile lengths. The tables also give data for 7.3- and 14.5-mile lengths, for combinations of pairs on different crossarms and for combinations of pairs on the second crossarm. As indicated on the table, some of the latter two types of pair-combinations were measured. A brief summary of conclusions to be drawn from these tables is given in the following.

Conclusions

1. The field trial indicates that the *REA-1* transposition system meets the objective that it retain much of the simplicity and economy of the *R1-5* transposition system and at the same time provides marked improvement in carrier frequency crosstalk losses.

2. From the standpoint of tolerable absorption loss and tolerable crosstalk losses, the maximum usable frequency is about 350 kc for a transposition pole spacing of 300 ft. The maximum usable frequency is inversely proportional to the transposition pole spacing.

3. Combinations of pairs on the top crossarm are in sufficient agreement with the crosstalk loss objectives up to 200 kc and for at least 29–30 miles. From 200–350 kc a crosstalk advantage (due to compandors, "FM capture," etc.) of 5–20 db is required depending on the top frequency. The crosstalk objectives were 50 db or more for far-end crosstalk loss and 40 db or more for near-end crosstalk loss.

4. Combinations of pairs involving one or more second crossarm pairs are in sufficient agreement with the crosstalk loss objectives up to 200 kc and for a length of 7–10 miles (assuming 300-ft transposition pole spacing). As expected, from 200 to a maximum of 300 kc, a crosstalk advantage of 5 to 20 db is required.

5. Further field tests of crosstalk losses up to 350 kc and for lengths of 30 miles or more might give additional data very useful for engineering purposes.

6. Crosstalk losses have not been carefully investigated for lengths longer than those given in Table I. A brief study indicates that doubling the maximum length given for any pair combination on this table would not usually cause any serious decrease in these losses insofar as they are controlled by the transposition patterns. Exceptional combinations are pair 9–10 to pair 13–14 and pair 3–4 to pair 11–12. For these combinations the far-end crosstalk losses are estimated to decrease about 6 db if the 14.5-mile length is doubled. The crosstalk losses, however, are determined not only by the transposition patterns but also by irregularities in construction. If it is desired to exceed the line lengths of Table I by an appreciable amount it is suggested that the rule for controlling transposition pole spacing irregularities shall be that the summed values of U^2 do not exceed $3L$ instead of $6L$.

It is hoped that, with the development of the *REA-1* transposition system, the authors and their associates have written a footnote to the history of better communications at a reasonable cost.

References

1. Application of the Type-P1 Carrier System to Rural Telephone Lines, **J. D. Howard, Jr.** *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 75, May 1956, pp. 205–14.

2. Telephone Lines for Rural Subscriber Service, **Lester Hockgraf, R. G. Watling.** *Ibid.*, vol. 74, May 1955, pp. 171–76.

3. Crosstalk on Open Wire Lines, **W. C. Babcock, E. Rentrop, C. S. Thaeler.** *Monograph no. 2520, Bell Technical Journal*, New York, N. Y., 1955.

4. Transposition System AE-150. *Bulletin 486*, Automatic Electric Company, Chicago, Ill., 1955.

5. A Transposition System for Carrier Systems up to 156 Kc, **Bess M. Kirkland.** *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 78, Nov. 1959, pp. 612–14.

6. K31 Rural Subscriber Carrier System. Kellogg Switchboard and Supply Company, Raleigh, N. C., sect. K31-KD, issue 1A, Jan. 1959.

7. K32 Open Wire Trunk Carrier System. Kellogg Switchboard and Supply Company, sect. K32-CNT, issue 1, May 1959.

8. A New 5-Channel Carrier System for Subscriber Use Employing Transistors, **B. R. Stachiewicz, R. L. Layburn, D. F. Jamieson.** *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 76, Sept. 1957, pp. 385–90.

9. Engineering Features and Field Trial Performance of a New Subscriber Carrier System, **R. L. Layburn.** *Ibid.*, Jan. 1958 section, pp. 681–87.

10. REA-1 Transposition System. *Telephone Engineering and Construction Manual—463*, Rural Electrification Administration, Washington, D. C., Jan. 1958.

11. Direct Capacitance and Capacitance Unbalance in Open Wire Telephone and Joint Use Lines. *Engineering Report no. 19*, Joint Subcommittee of Development and Research, Edison Electric Institute, New York, N. Y., and Bell Telephone System, New York, N. Y., Feb. 1933.

12. See reference 3, p. 101, table 17.

Analysis of Transcendental Nonlinear Systems

ALFRED A. WOLF
MEMBER AIEE

IN THE RECENT literature[1–6] certain aspects of a nonlinear theory were introduced which enable the solving of a class of nonlinear systems whose general dynamic equation is given by one like the following:

$$Z(D)x(t)+F\{x,\dot x,\ldots\}=g(t) \tag{1}$$

where the linear part $Z(D)x(t)$, the nonlinear part $F\{x,\dot x,\ldots\}$, and the driving function $g(t)$ are required to satisfy certain conditions given in references 1 and 2, and x is the response function. The dotted letters are time derivatives, so that $\dot x = d/dt$. The linear integro-differential operator has the form

$$Z(D)=\sum_{n=-N_1}^{N_2} f_n D^n$$

The coefficients f_n are constants for each value of n and the bounds N_1 and N_2 are positive, negative, or either one is zero, depending on $Z(D)$. Time-varying linear terms are easily added in a similar manner and offer no difficulty either of a theoretical or computational nature. The forcing function $g(t)$, although assumed here to be a deterministic process, may be relaxed to a random process with the aid of the transform ensemble theorem, as explained in reference 6.

In the papers cited,[1–6] the partition method was introduced based on the abovementioned restrictions. The method was applied to nonlinearities of the polynomial type. (The polynomials may include the function x and its derivatives in any complicated manner.) This paper shows how to apply the partition method to transcendental nonlinearities and illustrates the method with a simple example.

Solution of Differential Equations Containing Transcendental Nonlinearities

Many nonlinear problems in physics give rise to nonlinearities which may be represented as transcendents. For example, in solid-state physics, electronics, and transistor circuitry, equation 2 might be a representative equation describing the particular phenomena.[7]

$$Z(D)x(t)+be^x=g(t) \tag{2}$$

where $x=x(t)$ is the normalized solution. Another source of transcendence occurs in saturating systems.[8] Such systems as vacuum-tube circuitry, servomechanisms, etc., may be represented as

$$Z(D)x(t)+c\tanh x=g(t) \tag{3}$$

In synchronous machines[9] the stability of the machine, its pullout torque, etc., is usually studied by means of an equation of the following type:

$$J\ddot\theta+r\dot\theta+a\sin\theta=T \tag{4}$$

where

θ = angular displacement of the rotor
J = moment of inertia of the rotor
r = a resistive damping torque coefficient
$a\sin\theta$ = control torque
T = load torque

Under certain conditions it is possible to transform equations such as those given above into corresponding nonlinear differential equations of the polynomial type.[2] Then, when the resulting nonlinear differential equation is solved with algebraic types of nonlinearities, the solution for the differential equation with the transcendental nonlinearity is easily solved by reversing the transformation.

Solution Procedure

Given: A nonlinear differential equation such as

$$Z(D)x(t)+F(x)=g(t) \tag{5}$$

in which $F(x)$ is a transcendent. If the inverse function $F^{-1}(w)$ has an algebraic derivative, then equation 5 can be converted to a nonlinear differential equation with algebraic nonlinearities.

If the resulting algebraic nonlinearities satisfy the conditions previously imposed,[1,2] then a unique solution exists in the sense previously given. With this principle it is now possible to convert equations 2, 3, and 4 to the algebraic type of nonlinearity.

Consider, first, equation 2. Let

$$w=e^x \tag{6}$$

or its inverse

$$x=\log w \tag{7}$$

Then, on substituting equation 7 into equation 2 and differentiating w with respect to t in

$$Z(D)\log w+bw=g(t) \tag{8}$$

equation 8 becomes

$$Z(D)\frac{\dot w}{w}+bw=\dot g(t) \tag{9}$$

Let

$$Y(D)=\frac{1}{Z(D)} \tag{10}$$

and preoperate on each factor of equation 17 by $Y(D)$; simplifying yields

$$\dot w-bwY(D)\dot w-wY(D)\dot g=0 \tag{11}$$

In a similar manner equation 11 is converted to

$$\dot w+c(1-w^2)Y(D)\dot w=(1-w^2)Y(D)\dot g(t) \tag{12}$$

under the transformation

$$w=\tanh x \tag{13}$$

and equation 12 becomes

$$J\dot w+r\sqrt{1-w^2}\,\frac{1}{D}\,\frac{\dot w}{1-w^2}+ a\sqrt{1-w^2}\,\frac{1}{D}\,w=\sqrt{1-w^2}\,\frac{1}{D}\,T \tag{14}$$

under the transformation

$$w=\sin\theta \tag{15}$$

With the *partition method* described elsewhere[1,2] the transformed equations are easily solved.

Solution of a First-Order System

Consider a time-varying nonlinear transcendental system whose behavior is described by the following differential equation:

$$(1-t)\frac{dx}{dt}+e^x+t=0 \tag{16}$$

with boundary conditions

$$x(0)=0$$
$$\dot x(0)=-1 \tag{17}$$

Paper 60-879, recommended by the AIEE Basic Sciences Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting, Atlantic City, N. J., June 19–24, 1960. Manuscript submitted February 1, 1960; made available for printing April 26, 1960.

ALFRED A. WOLF is with the Stromberg-Carlson Company, Division of General Dynamics Corporation, Rochester, N. Y.

The author is deeply indebted to Dr. Y. H. Ku for his assistance and constant advice; and Dr. H. Sohon, Dr. J. G. Brainerd, Dr. R. Weller, and K. M. Lord for their generous and encouraging support. Grateful acknowledgment is made to the National Science Foundation for their support and encouragement in the preliminary aspects of this research, and to the Stromberg-Carlson Company for its support in continuing this work.

Following the artifice described, let

$$w = e^{x} \tag{18}$$

or

$$x = \log w \tag{19}$$

Substituting equation 19 into equation 16 yields

$$(1-t)\frac{\dot{w}}{w} + w = -t \tag{20}$$

Partitioning equation 20 at the $\dot{w}$ term gives

$$\dot{w} = -tw - w^{2} + t\dot{w} \tag{20A}$$

from which equation 21, and 22 the auxiliary equation, are obtained:

$$\dot{w} = \sum_{n=0}^{\infty} C_n t^{n} \tag{21}$$

$$\sum_{n=0}^{\infty} C_n t^{n} = -tw - w^{2} + t\dot{w} \tag{22}$$

Integrating equation 21 once,

$$w = \sum_{n=0}^{\infty} \frac{C_n}{n+1} t^{n-1} + K \tag{23}$$

where $K=1$, in order to satisfy the given boundary conditions. Combining equation 23 with equation 22 yields the recurrence relation

$$C_n = -\frac{C_{n-2}}{n-1} - C_{n-1}{}^{(0)} - C_{n-2,0}{}^{(2)} - \frac{2C_{n-1}}{n} - C_n{}^{(0)} + C_{n-1} \tag{24}$$

Solving the recurrence relation by successive substitution yields:

For $n=0$:

$$C_0 = -1 \tag{25}$$

For $n=1$:

$$C_1 = -1 - 2C_0 + C_0 = 0 \tag{26}$$

For $n=2$:

$$C_2 = -C_0 - C_{0,0}{}^{(2)} - 2\frac{C_1}{2} + C_1 \tag{27}$$

Since

$$C_{n-2,0}{}^{(2)} = \sum_{k=0}^{n-2} \frac{C_k}{k+1}\frac{C_{n-k-2}}{n-k-1} \tag{28}$$

then

$$C_{0,0}{}^{(2)} = C_0{}^{2} \tag{29}$$

Hence

$$C_2 = -C_0 - C_0{}^{2} = 1 - 1 = 0 \tag{30}$$

For $n=3$

$$C_n = -\frac{C_{n-2}}{n-1} - \sum_{k=0}^{n-2} -\frac{C_k C_{n-k-2}}{(k+1)(n-k-1)} - \frac{2C_{n-1}}{n} + C_{n-1} \tag{31}$$

Since each term of the right member will depend on C_1 and C_2 for $n=3$, then $C_3=0$, and since successively higher terms will depend on, first, C_2, C_3 which are zero, then C_3, C_4, which are also zero, it may be concluded that

$$C_n = 0 \text{ for all } n \geq 3 \tag{32}$$

Substituting the values of the C's as determined above into equation 23 gives

$$w = 1 - t \tag{33}$$

Applying the inverse transformation, equation 19, gives the solution of equation 16 as

$$x(t) = \log(1-t) \tag{34}$$

which satisfies the given boundary conditions.

Limitations and Possibility of Extension

The scheme presented in this paper somewhat extends the method of partition to nonlinear functions other than algebraic types. A limitation of this scheme occurs when the resulting derivative gives rise to nonlinear functions which are not consistent with the required conditions. In some such cases special treatment can be applied to the resulting nonlinearity so as to bring it into consistency with the required conditions. An example of this occurs in equations like equation 4. We see that equation 4 goes into equation 14, which has fractional nonlinearities. These must be treated with special care.

The modified forcing function $A(t)$ which is given in equation 35 by partitioning equation 1 at, say, the linear part, may be expanded in a number of ways:

$$Z(D)x(t) = A(t) \tag{35}$$

One such expansion is a Dirichlet series.[10,11] This is proved in reference 2. Combining this expansion of $A(t)$ with the technique described here increases the power of the method since it is possible to examine periodic and exponential behavior easily.

It is also to be noted that all but the highest-order derivative may be transposed to the right member. In this case the moments of the folded impulse response $y(t)$ given below are power functions. By the moment theorem[1,2] the solution of equation 1, which is the linear combination of all the moments of the folded impulse response of the linear part, is in this case a power series. If the partition is made so as to leave the highest-order derivative, say the kth derivative, and the next lower order derivative, say the $(k$-$1)$ derivative in the left member, then the solution will consist of sums of entire functions of exponential type of order 1 that are nonperiodic. If the kth derivative and the $(k-1)$ derivative remain in the left member after partitioning, the solution will consist of sums of entire functions[12,13] of exponential type of order 1 that are periodic. A detailed development of the partition process and its effect on the nature of the solution may be the subject of a future work.

Conclusions

This paper has presented a simple means of extending the partition theory to the case of nonlinear functions that are transcendental. It was shown that whenever the transcendent is such that its inverse has an algebraic derivative, the differential equation can be transformed into one that contained only algebraic nonlinearities. The solution then followed the method prescribed in earlier work.[1–6] The recurrence relations can be formulated as described in references 1 and 2 or by the Taylor-Cauchy transform given in reference 5.

The method of transformation of variables presented here fails in the case of nonlinear functions consisting of algebraic and transcendental functions. To solve such problems other methods must be applied. An approximate solution to this kind of problem was developed in reference 8. An exact solution of this problem in the sense represented by references 1 and 2 has not yet been determined.

Although the forcing function is assumed to be of an exponential type, it is possible to relax this condition, as described in reference 6, to random functions. It is then possible, with the procedure given here with the Taylor-Cauchy transform, to study, among other things, the effects of saturation on random inputs.

Appendix I. Restrictions

Restrictions are as follows:

1. The forcing function, $g(t)$ is of exponential type, and order 1.
2. The impulse response $y(t)$ given below is of exponential type:

$$y(t) = \frac{1}{2\pi j}\int_{Br} \frac{e^{st}}{Z(s)}\, ds$$

3. The nonlinear part $F(x)$ is a continuous, single-valued function of x; moreover, it must satisfy the following condition:

Given $M>0$ and $a>0$, and two continuous functions $u(t)$ and $v(t)$ which are asymptotically like e^{-at}, then

$$|F(u) - F(v)| < Me^{-at}|u - v|$$

This is a modified Lipshitz condition.[1,2,14]

Appendix II. Summary of Principle of Partition

The principle of partition consists essentially of the following steps satisfying certain mentioned conditions:

1. The dynamic equation representing the system is partitioned so that the linear term with the highest-order derivative are contained in one member of the equation.
2. A modified forcing function $A(t)$ is then obtained which consists of the actual forcing function, nonlinear and/or some linear terms (but not the highest-order derivative).
3. A proper choice is made of the expansion of $A(t)$ according to the particular properties of the linear, nonlinear, and driving functions. An auxiliary equation is thus formed.
4. The properly partitioned differential equation is then solved.
5. Certain coefficients which arise from the auxiliary equation are then calculated.
6. These coefficients which appear in the partitioned equation, are then eliminated to obtain the exact solution.

References

1. Recurrence Relations in the Solution of a Certain Class of Nonlinear Systems, A. A. Wolf. *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 78, 1959 (Jan. 1960 section), pp. 830–34.

2. A Mathematical Theory for the Analysis of a Class of Nonlinear Systems, A. A. Wolf. *Ph.D. Dissertation*, The Moore School of Electrical Engineering, University of Pennsylvania, Philadelphia, Pa., June 1958.

3. A Stability Criterion for Nonlinear Systems, Y. H. Ku, A. A. Wolf. *AIEE Transactions*, pt. II (*Applications and Industry*), vol. 78, July 1959, pp. 144–48.

4. On the Significance of Transients and Steady State Behavior in Nonlinear Systems, A. A. Wolf. *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 47, no. 10, Oct. 1959, pp. 1785–86.

5. Taylor-Cauchy Transforms for Solving a Certain Class of Nonlinear Systems, Y. H. Ku, A. A. Wolf, J. H. Dietz. *IRE Convention Record*, Institute of Radio Engineers, pt. 2, Circuit Theory, 1959, pp. 49–61; also *Proceedings* (in modified form), vol. 48, no. 5, May 1960, pp. 912–22.

6. Transform-Ensemble Method for Analysis of Linear and Nonlinear Systems with Random Inputs, Y. H. Ku, A. A. Wolf. *Proceedings*, National Electronics Conference, Chicago, Ill., vol. 15, 1960, pp. 441–55.

7. Crystal Rectifiers (book), H. C. Torrey, C. A. Whitmer. McGraw-Hill Book Company, Inc., New York, N. Y., vol. 15, 1948.

8. On a Systematic Approximation to the Partition Method for Analysis of a Class of Nonlinear Systems, Y. H. Ku, A. A. Wolf, J. H. Dietz. *AIEE Transactions*, pt. II (*Applications and Industry*), vol. 79, 1960 (*Paper no. 60*-111).

9. Nonlinear Vibrations (book), J. J. Stoker. Interscience Publishers, New York, N. Y., 1950.

10. The Taylor Series (book), P. Dienes. Dover Publications, Inc., New York, N. Y., 1957.

11. Theory of Functions (book), E. C. Titchmarsh. Oxford University Press, London, England, 1931.

12. Entire Functions (book), R. P. Boas. Academic Press, New York, N. Y., 1954.

13. Theorie und Anwendung der Laplace Transformation (book), G. Doetsch. Dover Publications, Inc., 1943.

14. Existence Theorems (book), F. J. Murray, K. S. Miller. New York University Press, New York, N. Y., 1954.

Theory of 3-Phase Bridge Magnetic Amplifier

II—The Time Constant

HERBERT F. STORM
MEMBER AIEE

CLARENCE W. FLAIRTY
ASSOCIATE MEMBER AIEE

THE OBJECT of this paper is to derive the time constant of a 3-phase bridge-connected magnetic amplifier from its static control characteristic. The bridge connection is the most widely used 3-phase connection because of its high degree of utilization for rectifiers and magnetic components; it was therefore chosen for this study.

The time constant is derived from the static control characteristic of the amplifier, and the dynamic response is presented in the form of a transfer function. Test data and calculated data are compared.

A prerequisite for studying the dynamics of a 3-phase bridge amplifier is an understanding of its steady-state properties. The latter properties are described in detail in Part I of this paper;[1] for further material on 3-phase amplifiers see references 2 through 16.

Circuit Connections and Supply System

The circuit under consideration is shown in Fig. 1(A); it is called a 3-phase bridge magnetic amplifier. The amplifier consists of six identical saturable reactors $A, C' \ldots, B'$, and six identical rectifiers rec A, rec $C' \ldots$, rec B'. The load is a pure resistance.

The terminals 1, 2, and 3 are connected to a symmetrical 3-phase system, having the star voltages

$$e_{1-0}=\frac{E_m}{\sqrt{3}}\cos\theta \tag{1A}$$

$$e_{2-0}=\frac{E_m}{\sqrt{3}}\cos\left(\theta-\frac{2\pi}{3}\right) \tag{1B}$$

$$e_{3-0}=\frac{E_m}{\sqrt{3}}\cos\left(\theta-\frac{4\pi}{3}\right) \tag{1C}$$

with

$$\theta=\omega t \tag{2}$$

and

$$\omega=2\pi f \tag{3}$$

The line-to-line voltages e_{1-2}, e_{2-3}, e_{3-1} are defined by equations

$$e_{1-2}=e_{1-0}-e_{2-0} \tag{4A}$$

$$e_{2-3}=e_{2-0}-e_{3-0} \tag{4B}$$

$$e_{3-1}=e_{3-0}-e_{1-0} \tag{4C}$$

and, when substituted in equations 1, they result in

$$e_{1-2}=E_m\sin\left(\theta+\frac{2\pi}{3}\right) \tag{5A}$$

$$e_{2-3}=E_m\sin\theta \tag{5B}$$

$$e_{3-1}=E_m\sin\left(\theta-\frac{2\pi}{3}\right) \tag{5C}$$

The supply voltages are shown in Fig. 1(B) in phasor form and in Fig. 1(C) in sinusoidal form. With the use of the phasor diagram the instantaneous values of the voltages are obtained by rotating the entire phasor diagram counterclockwise and by taking the projections of the phasors on the stationary base line, indicated by $\theta=0$.

The dots in Fig. 1 signify winding polarities. The voltage across a winding is defined as positive when the terminal with the dot is more positive than the other terminal. A current is called positive when it flows in the direction of the arrow. The magnetomotive force is called positive when derived from a positive current and the core fluxes are called positive

---

Paper 60-872, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting, Atlantic City, N. J., June 19–24, 1960. Manuscript submitted February 8, 1960; made available for printing April 15, 1960.

Herbert F. Storm and Clarence W. Flairty are with the General Electric Company, Schenectady, N. Y.

l, 79
1
60

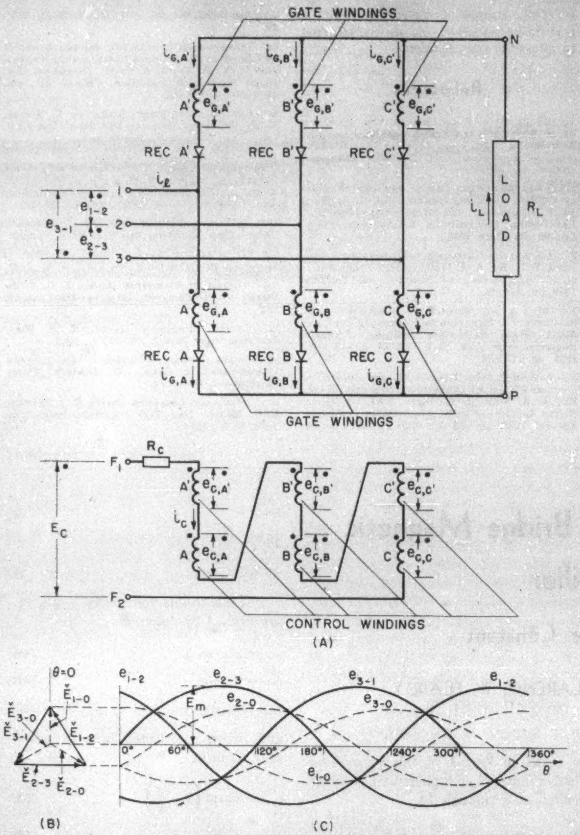


Fig. 1. Three-phase bridge amplifier. A—Circuit diagram. B—Phasor diagram of supply voltages. C—Sinusoidal representation of supply voltages

when derived from positive magnetomotive forces.

Control Circuit

The resistances of the control windings and any external circuit resistance are lumped into the control circuit resistance R_C; see Fig. 1(A). Then, the equation for the control circuit is

$$E_C=e_{C,A}+e_{C,C'}+e_{C,B}+e_{C,A'}+e_{C,C}+e_{C,B'}+i_CR_C \quad (6)$$

If we consider magnetic flux due to ferromagnetism only, the above voltages can be expressed in terms of the core fluxes of the saturable reactors:

$$e_{C,A}=\frac{d\phi_A}{dt}\,\frac{N_C}{10^8} \quad (7A)$$

$$e_{C,C'}=\frac{d\phi_{C'}}{dt}\,\frac{N_C}{10^8} \quad (7B)$$

$$\vdots$$

$$e_{C,B'}=\frac{d\phi_{B'}}{dt}\,\frac{N_C}{10^8} \quad (7F)$$

Denoting the total flux linkage as seen by the control circuit by

$$\psi_C=(\phi_A+\phi_{C'}+\phi_B+\phi_{A'}+\phi_C+\phi_{B'})N_C \quad (8)$$

yields from equations 6 through 8

$$E_C=i_CR_C+\frac{d\psi_C}{dt}\,10^{-8} \quad (9)$$

Equation 9 bears considerable similarity to the differential equation of a linear reactor circuit. For instance, in both cases and with zero circuit resistance, the total flux linkage increases linearly with time, if the control voltage E_C is a pure d-c quantity and remains constant. One difference between the two circuits consists in the presence of ripple in the case of the magnetic amplifier. This ripple has a fundamental frequency of $6f$, and higher harmonics. The ripple not only pervades the core flux, but also practically all other electrical quantities occurring in the amplifier.

For the purpose of further analysis, the various quantities will be separated into an average component and an alternating component. For instance, the flux linkage ψ_C can be written

$$\psi_C=\Psi_C+\psi_{C,a-c} \quad (10A)$$

where Ψ_C is the average component

$$\Psi_C=\frac{3}{\pi}\int_0^{\pi/3}\psi_C\,d\theta \quad (10B)$$

Similarly, the control current i_C is

$$i_C=I_C+i_{C,a-c} \quad (10C)$$

where I_C is the average component

$$I_C=\frac{3}{\pi}\int_0^{\pi/3}i_C\,d\theta \quad (10D)$$

When, from equation 10A, c is substituted into equation 9, the five terms can be grouped into E_C, I_C, Ψ_C, and $i_{C,a-c}$, $\psi_{C,a-c}$. If we stipulate that this analysis shall be applied to transients whose duration is long by comparison to the duration of an averaging cycle ($\Delta t=1/6f$), the a-c components cancel each other, resulting in

$$E_C=I_CR_C+\frac{\Delta\Psi_C}{\Delta t}\,10^{-8} \quad (11)$$

The differential quotient in equation 9 was replaced by a difference quotient because of the sampled character of the average quantities.

Assumptions

The following assumptions will be made: The duration of the transient is long by comparison with the duration of $1/6f$. The saturation branches of the hysteresis loops of the reactor cores are horizontal. Fluxes due only to ferromagnetism are considered. At the so-called minimum load point M in Fig. 2 the core fluxes swing just from negative saturation to positive saturation (normal excitation). The rectifier reverse resistance is infinity. The resistance of the control circuit is small by comparison with the (unsaturated)

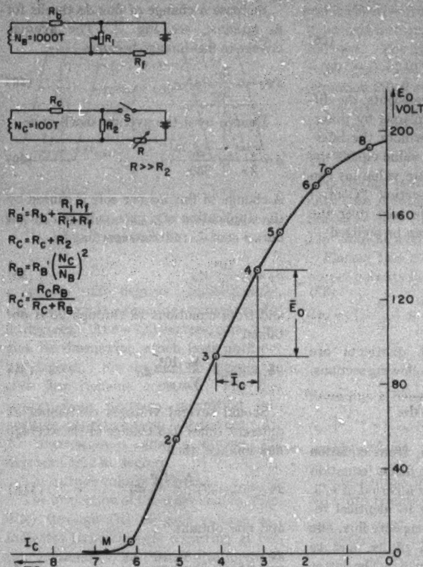


Fig. 2. Control characteristic of amplifier. Numbers on curve representing characteristic refer to measurements of transients in Table I

magnetizing reactance formed by the series-connected control windings. Gate resistance R_G and rectifier forward resistances R_f are lumped externally into a resistance in series with the load resistance R_L; the total of these resistances form the output circuit, and its value R_0 is called the output resistance:

$$R_0 = R_L + 2R_G + 2R_f \tag{12}$$

The contributions of the magnetizing currents of the saturable reactors to the magnitude of the current flowing in the output circuit will be neglected.

A magnetic amplifier such as shown in Fig. 1 can be viewed as a switching device which will admit the supply voltage to the load resistance for a controllable fraction of the period of the supply voltage. The supply voltage when applied to the output circuit is called the output voltage e_0. Hence, the load current i_L becomes

$$i_L = \frac{e_0}{R_0} \tag{13A}$$

The average E_0,

$$E_0 = \frac{3}{\pi}\int_0^{\pi/3} e_0 d\theta \tag{13B}$$

and the average load current I_L,

$$I_L = \frac{E_0}{R_0} \tag{13C}$$

In calculating the performance of feedback-control systems by means of Bode or Nyquist diagrams, only the changes of the variables called the incremental variables are of interest. The following analysis will deal mostly with incremental variables, designated by a bar over the symbol. Where the transient response has to be finally expressed in absolute values, the law of superposition is applied: The instantaneous absolute value of the variable during the transient equals the variable before the transient, plus the incremental variable during the transient.

In terms of incremental values, equation 11 becomes

$$E_C = \bar{I}_C R_C + \frac{\Delta\Psi_C}{\Delta t} 10^{-8} \tag{14}$$

Derivation of Response

The first term of equation 14 can be expressed by

$$E_C = I_C R_C \tag{15}$$

Defining the incremental voltage gain K_E of the amplifier by

$$K_E = \frac{E_0}{E_C} \tag{16}$$

and expanding the second term in equation 14 yields

$$E_C = \frac{E_0}{K_E} + \frac{\Delta\Psi_C}{\Delta E_0}\frac{\Delta E_0}{\Delta t} 10^{-8} \tag{17}$$

Based on the assumption of long transients, the difference quotient $\Delta E_0/\Delta t$ is approximated by the differential quotient dE_0/dt. With Heaviside notation

$$\frac{dE_0}{dt} = pE_0 \tag{18}$$

To accentuate the similarities in dynamic response between single-phase and 3-phase bridge amplifiers, the factor m is introduced:

$$m = \frac{\Delta\bar{\Psi}_C}{\Delta E_0} 2f \frac{N_G}{N_C} 10^{-8} \tag{19}$$

Substituting equations 18 and 19 in equation 14 yields

$$E_C = \frac{E_0}{K_E} + pE_0 \frac{1}{2f}\frac{N_C}{N_G} m \tag{20}$$

For incremental variations of the control voltage E_C, the voltage gain K_E and the factor m can be considered constant, and one obtains as the solution of equation 20:

$$E_0 = \frac{E_C K_E}{1 + p\frac{1}{2f} K_E \frac{N_C}{N_G} m} \tag{21}$$

Equation 21 signifies an exponential response with the so-called control circuit time constant T_C

$$T_C = \frac{1}{2f} K_E \frac{N_C}{N_G} m \tag{22}$$

With a single-phase bridge magnetic amplifier, $m = 1$ (see reference 4, p. 145).

If no other circuits, such as bias circuits, are coupled to the control windings, the total time constant T of the magnetic amplifier is equal to the control circuit time constant T_C. With other resistive circuits coupled to the control windings,

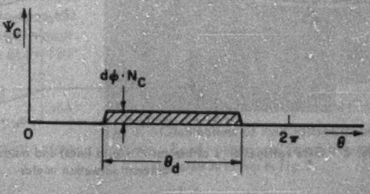


Fig. 3. Illustration for calculation of average flux linkages; see equation 28

T consists of the sum of T_C and the other time constants, explained in reference 4, p. 161.

From equations 21 and 22 one obtains for the transient response of the output voltage

$$E_0 = K_E E_C \frac{1}{1+pT_C} \tag{23}$$

The transient response of I_L becomes (equation 13C)

$$I_L = \frac{E_C K_E}{R_0} \frac{1}{1+pT_C} \tag{24}$$

and the response of the control current I_C becomes

$$I_C = \frac{E_C}{R_C} \frac{1}{1+pT_C} \tag{25}$$

For the qualitative determination of the transient response, K_E (equation 16) and m (equation 19) need to be known. The incremental voltage gain can be determined from a static control characteristic such as shown in Fig. 2, with the use of equations 15 and 16; m is determined analytically. To do this, the saturation angle α is parametrically introduced into the difference quotient of equation 19. To simplify the notations the saturation angle will be called α (not α_s as in part I of the paper).

$$\frac{\Delta\bar{\Psi}_C}{\Delta\bar{E}_0} = \frac{\dfrac{\Delta\bar{\Psi}_C}{\Delta\bar{\alpha}}}{\dfrac{\Delta\bar{E}_0}{\Delta\bar{\alpha}}} \tag{26}$$

As we consider long transients, the difference quotients are replaced by differential quotients. Furthermore, the differential of an incremental value equals the differential of the absolute value (as can be deduced from the section "Control Circuit") and hence, the bars over the symbols of differentials can be omitted:

$$m = \frac{\dfrac{d\Psi_C}{d\alpha}}{\dfrac{dE_0}{d\alpha}} 2f \frac{N_G}{N_C} 10^{-8} \tag{27}$$

How these differential quotients are obtained is shown in the following sections.

Determination of $d\Psi_C/d\alpha$

To determine $d\Psi_C/d\alpha$, from equation 27, the sum of all six core fluxes (equation 8) may be averaged over a period of $\pi/3$. Another method leading to identical results is to consider only one core flux, but averaged over a period of 2π, and to multiply this average by 6. The latter method is more practical, and will be subsequently used.

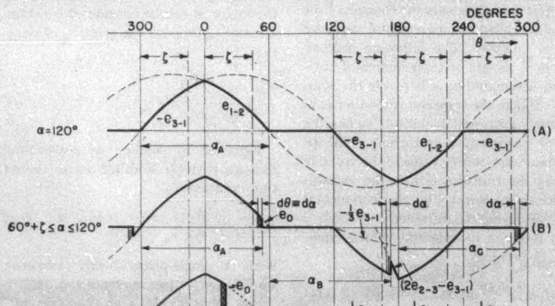


Fig. 4. Gate voltages $e_{G,A}$ of reactor A (solid lines) and output voltages e_0 (dashed lines) for different saturation angles

Suppose a change of flux $d\phi$ dwells for θ_d radians; see Fig. 3. The average change in flux linkage becomes

$$d\Psi_C = 6 \frac{\theta_d}{2\pi} d\phi N_C \tag{28}$$

Denote by ϑ the average dwell of $d\phi$

$$\vartheta = \frac{\theta_d^{\text{rad}}}{2\pi} \text{ or } \frac{\theta_d^{\circ}}{360} \tag{29}$$

A change of flux $d\phi$ per core is caused by the application of a gate voltage e_G for a duration of $d\alpha$ radians according to

$$d\phi = e_G \frac{10^8}{\omega N_G} d\alpha \tag{30}$$

and from equations 28 through 30 is obtained

$$d\Psi_C = 6(e_G\vartheta) \frac{N_C}{N_G} \frac{10^8}{\omega} d\alpha \tag{31}$$

Should several voltages contribute at different times to a change of the average flux linkage, then

$$d\Psi_C = 6\Sigma(e_G\vartheta) \frac{N_C}{N_G} \frac{10^8}{\omega} d\alpha \tag{32A}$$

and one obtains

$$\frac{d\Psi_C}{d\alpha} = 6\Sigma(e_G\vartheta) \frac{N_C}{N_G} \frac{10^8}{\omega} \tag{32B}$$

The gate voltages e_G have been determined;[1] hence equation 32B can be evaluated. The gate voltages are shown in Fig. 4 for the gate voltage $e_{G,A}$ of reactor A. The voltages of the other gates have the same wave shape, but are phase-displaced sequentially by 60 degrees, following in the order of A, C', B, A', C, B'.

The assumptions under which these wave shapes were obtained are the same as those outlined in the section "Assumptions." More specifically, the hysteresis loop of the saturable reactor is shown in Fig. 5(A). The slopes of the dynamic loop may vary from zero to infinity, and no restrictions are placed on the width of the d-c hysteresis loop. The hysteresis loop may also have the form shown in Fig. 5(B), provided

$$\frac{AT_{c,dc}}{AT_{c,0}} \geq 1 - 2(\sqrt{3}-1) \tag{33}$$

Determination of Factor m, Equation 27

Minimum Load Point $\alpha = 120$ Degrees. The gate voltage $e_{G,A}$ for saturable reactor A is shown in Fig. 4(A). The saturation angle α is counted from the instant when the flux begins its upswing until saturation is reached. The upswing of the flux in core A begins at $\theta = 300$ degrees and ends at $\theta = 60$ degrees; hence

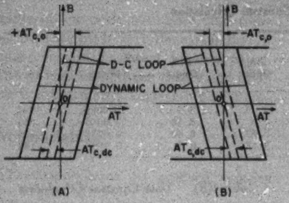


Fig. 5. Dynamic and d-c B-H loops; see equation 33

α equals 120 degrees. Subsequently, the core flux remains constant for 60 degrees. At $\theta = 120$ degrees, core A begins its downswing which lasts until $\theta = 240$ degrees. For the next 60 degrees the core flux remains constant. The flux swings from negative saturation into positive saturation from 300 degrees $\leq \theta \leq$ 60 degrees, and returns during 120 degrees $\leq \theta \leq 240$ degrees.

The output voltage e_0 is zero.

The inspection of the wave shapes, Figs. 4(A) through (E) of gate voltage $e_{G,A}$ indicates basic changes occurring at certain boundaries, and hence the factor m is determined separately for the various ranges.

From Fig. 4(B) $60° + \zeta \leq \alpha \leq 120°$. Suppose the saturation of core A occurs at $\theta = -60° + \alpha_A$; at this instant the gate voltage $e_{G,A}$ vanishes from A and reappears in the output circuit as the output voltage e_0, shown by a dashed line in Fig. 4(B). At $\theta = 60° + \alpha_B$, core B saturates and, as a result, the gate voltage $e_{G,A}$ drops to $(2e_{2-3} - e_{3-1})$. At $\theta = 180° + \alpha_C$, core C saturates and $e_{G,A}$ reappears. The angles α_A, α_B, and α_C are identical in magnitude, and different indexes merely point to that core which goes into saturation.

Let us now introduce a change of control voltage in the positive direction. As a result, corresponding changes follow for the saturation angle, control current, and gate and output voltages. The voltage changes are shown in Fig. 4 by various forms of shading: the changes in output voltage by crosshatching, the changes in gate voltage by vertical or horizontal shading, for positive and negative differential contributions respectively. The contributions of the vertically or horizontally shaded gate voltages produce the change of flux linkages according to equation 32A. In the case shown in Fig. 4(B) there are two contributions to change the flux linkage; The first occurs at $\theta = 60° + \alpha$; the contributing voltage is

$$e_{G,A} = -e_{3-1} - (2e_{2-3} - e_{3-1}) = -2e_{2-3} \quad \text{(34A)}$$

and is effective from $\theta = 60° + \alpha$, until core A returns into positive saturation at $\theta = 300° + \alpha$. The dwell angle $\theta_d = (300° + \alpha) - (60° + \alpha) = 240°$, and the average dwell (equation 29) $\vartheta = 2/3$.

The second contribution occurs at $\theta = 180° + \alpha$. The contributing voltage is

$$e_{G,A} = -e_{3-1} \quad \text{(34B)}$$

This voltage is effective from $\theta = 180° + \alpha$ until core A returns into positive saturation at $\theta = 300° + \alpha$. The dwell angle $\theta_d = 120°$ and the average dwell $\vartheta = 1/3$.

Placing this contributing voltage with proper polarity into equation 32A yields

$$d\Psi_C = 6\left(-\frac{4}{3}\,e_{2-3}\Big|^{60°+\alpha} + \frac{1}{3}\,e_{3-1}\Big|^{180°+\alpha}\right) \times \frac{N_C}{N_G}\,\frac{10^8}{\omega}\,d\alpha \quad \text{(35A)}$$

From Fig. 1(B) and (C)

$$e_{2-3}\Big|^{60°+\alpha} = -e_{3-1}\Big|^{\alpha} \quad \text{(34C)}$$

$$e_{3-1}\Big|^{180°+\alpha} = -e_{3-1}\Big|^{\alpha} \quad \text{(34D)}$$

Substituting from equations 34C and 34D into equation 35A yields

$$\frac{d\Psi_C}{d\alpha} = 6e_{3-1}\Big|^{\alpha}\,\frac{N_C}{N_G}\,\frac{10^8}{\omega} \quad \text{(35B)}$$

and the numerator of equation 27 is thus obtained.

The denominator of equation 27 is the differential of the output voltage dE_0. As shown in Fig. 4(B) the output voltage e_0 equals voltage e_{1-2} from $\theta = -60° + \alpha$ to $\theta = 60°$. Hence

$$dE_0 = 6\,\frac{1}{2\pi}\,(-e_{1-2})\Big|^{-60°+\alpha} d\alpha \quad \text{(35C)}$$

From Fig. 1(B) and (C)

$$-e_{1-2}\Big|^{60°+\alpha} = e_{3-1}\Big|^{\alpha} \quad \text{(34E)}$$

Substituting from equation 34E in equation 35C yields

$$\frac{dE_0}{d\alpha} = \frac{3}{\pi}\,e_{3-1}\Big|^{\alpha} \quad \text{(35D)}$$

Applying equations 35B and 35D to equation 27 yields, finally, for m

$$m = 2 \quad \text{(35E)}$$

As the control current is increased in the positive direction, α continues to decline, and one end point of the voltage contribution follows along the voltage $(2e_{2-3} - e_{3-1})$ until this voltage intersects with voltage $-1/3\ e_{3-1}$, whereupon the end point follows along the latter voltage. The above intersection occurs at $\theta = 120° + \zeta$; hence ζ follows from

$$-\frac{1}{3}\,e_{3-1} = 2e_{2-3} - e_{3-1} \quad \text{(36A)}$$

Simplifying and substituting into equation 5 results in

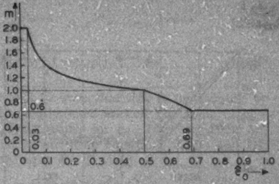


Fig. 6. Factor m, for equation 22, as function of per-unit output voltage $\hat{e}_0$

$$\zeta = 46.1° \quad \text{(36B)}$$

One concludes that $m = 2$ is valid only for saturation angles between 106.1° and 120°. The angle expressed by equation 36B was termed θ_u in Part I of this paper.

Sometimes it is more illustrative to describe the validity range for m in terms of the per-unit output voltage $\hat{e}_0$:

$$\hat{e}_0 = \frac{E_0}{E_{0,m}} \quad \text{(37A)}$$

where E_0 is the average value of the output voltage, and $E_{0,m}$ its maximum average value (at full saturation). This $\hat{e}_0$ is obtainable from Part I of the paper.

$$\hat{e}_0 = 1 - \sin(\alpha - 30°) \quad \text{(37B)}$$
$$60° \leq \alpha \leq 120°$$

$$\hat{e}_0 = \cos\alpha \quad \text{(37C)}$$
$$0 \leq \alpha \leq 60°$$

For the angular range $60° + \zeta \leq \alpha \leq 120°$, one obtains from equations 36B and 37B the validity range for m in terms of per unit output voltage $0 \leq \hat{e}_0 \leq 0.03$.

From Fig. 4(C), $60° \leq \alpha \leq 60° + \zeta$. In this case the flux linkage is affected by three contributions, as outlined in Appendix I, resulting in

$$m = 1 + \frac{1}{3}\,\frac{\sin(\alpha - 60°)}{\sin(120° - \alpha)} \quad \text{(38E)}$$

The validity range for m in terms of $\hat{e}_0$ follows from equation 37B: $0.03 \leq \hat{e}_0 \leq 0.5$.

From Fig. 4(D), $\zeta \leq \alpha \leq 60°$. As shown in Appendix II,

Fig. 7. Oscillograms of load voltage during transient. Load voltage was varied between points as shown in Fig. 2

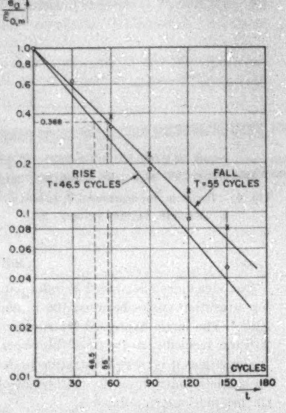


Fig. 8. Transients of Fig. 7 plotted on semilog paper

$$m = 1 - \frac{\sin(60° - \alpha)}{\sin 60°} \qquad (39F)$$

The validity range for equation 39 follows from equation 37C: $0.5 \le \bar{e}_0 \le 0.693$.

From Fig. 4(E), $\theta \le \alpha \le \zeta$. In Appendix III,

$$m = \frac{2}{3} \qquad (40B)$$

The validity range is $0.693 \le \bar{e}_0 \le 1.0$.

To facilitate the calculation of the time constant (equation 22), m was evaluated for various saturation angles (equations 35E, 38E, 39F, and 40B), and plotted versus $\bar{e}_0$ (equations 37B, and 37C) in Fig. 6.

For single-phase bridge amplifiers m is constant (equaling unity). It follows that for the single-phase amplifier the output voltage E_0 and the flux linkages Ψ_C are proportional throughout the entire control range. In contrast, with 3-phase bridge amplifiers m varies from 2 to 2/3; hence, the output voltage and the flux linkages are no longer proportional throughout the entire control range. Prior to this theoretical analysis, this difference between single-phase and 3-phase bridge amplifiers was discovered experimentally by L. C. Harriott and F. J. Ellert of the General Electric Company.

Tests

Tests were conducted on an amplifier connected as shown in Fig. 1. The cores are of toroidal construction, with the use of 4-mil Deltamax tape. The rectifiers are of the germanium type *G.E. 4JA211 CH1AC5*. The maximum power output of the amplifier is 1 kw, with a supply frequency of 60 cycles per second.

Table I(A). Time Constant Calculation

Measured Span	ΔE_0, Volts	ΔI_C, Ma	R_C, Ohms	R_B, Ohms	R_C', Ohms	$\Delta I_C R_C'$, Volts	$\frac{\Delta E_0}{\Delta I_C R_C'}$
1-2	49.6	11.3	34.7	7.11	5.9	0.0666	745
3-4	40.1	10	34.7	7.5	6.16	0.0616	651
5-6	24.2	11.3	34.7	7.87	6.42	0.0725	334
7-8	11.8	11.6	34.7	8.13	6.6	0.0765	154

As the theory is based on long transients, the first checks were aimed at this category. The saw-tooth wave shape of the output voltage makes its evaluation tedious when average values are sought. To simplify the measurement of the average values, a low-pass filter was connected to the output of the amplifier. Such a low-pass filter, however, has the drawback of modifying the transient response as measured. By keeping the filter time constant small with respect to the time constant of the amplifier, the effect of the filter time constant appears only at the beginning of the transient. Two transients of the output voltage of the amplifier including the filter are shown in Fig. 7. One transient is for rising output voltage, the other for falling output voltage. To determine the character of these transients the voltage ratios $e_0/E_{0,m}$ are plotted on semilog paper, resulting in the points marked in Fig. 8; $E_{0,m}$ is the maximum steady-state change of the load voltage, and e_0 signifies the change of load voltage that is necessary to reach $E_{0,m}$.

For true exponential response, the points should lie on a straight line out of the origin. Actually, the points near the origin form a curve, which straightens out as time increases. By drawing through the origin a line parallel to the straight section of the response (Fig. 8), and intersecting it at $e_0/E_{0,m} = 1/\epsilon = 0.368$, two time constants are obtained: for rise, $T = 46.5$ cycles; for fall, $T = 55$ cycles. The fall time constant is somewhat higher because the control voltage can induce voltages in the gate windings which are in the forward direction of the rectifiers and thus may create coupled circuits. In this case, the time constants of the coupled circuits add to the control circuit time constant T_C.

In the calculation of the time constant according to equation 22, and shown in Table I, no coupled circuits through rectifiers were considered, and therefore the calculated values should be compared with the measured data for rise time constants. The time constants in Table I are expressed in cycles, and are obtainable from equation 22 by omitting the frequency f.

Table I(B). Time Constant Calculation

Measured Span	$\frac{\Delta E_0 N_G}{\Delta I_G R_G' N_C}$	m	Time Constant, Cycles: Transient Increasing, Calculated	Time Constant, Cycles: Transient Increasing, Measured	Time Constant, Cycles: Transient Decreasing, Measured
1-2	111	1.3	72	78	
3-4	97	0.9	43.6	46.5	55
5-6	49.8	2/3	16.7	18	26
7-8	23	2/3	7.7	10	14

The ranges over which the time constants of Table I were measured are indicated by numerals in Fig. 2. The insert of Fig. 2 shows the control arrangement, where a bias winding was used to establish one starting point of the transient, and the control winding was used to produce the change in control voltage. Table I is presented not only to provide a comparison between theory and practice, but also to indicate the steps necessary for the calculation of the time constant.

Tests were also conducted to check the validity of equation 22 for relatively smaller time constants, a sample being shown in Table II. In measuring small time constants, a low-pass filter in the output circuit is no longer practical for obtaining average values. Instead, it becomes necessary to determine geometrically the voltage areas in the oscillograms, a procedure not characterized by high precision. In addition, the dead-time effect (not analyzed in this paper) requires consideration.

Conclusions

The time constant of a 3-phase magnetic amplifier can be readily determined from its static control characteristic. Measured data are shown for time con-

Table II. Smaller Time Constant Calculation

Per-Unit Output Voltage	Transient Increasing: Calculated	Transient Increasing: Measured	Transient Decreasing, Measured
0.333	7.6 3.8	7.5 3.9	8.9 5.1
0.61	0.83	1.15	

stants of one cycle and more, and comparisons are made with the calculation of time constant proposed in this paper. The correlation between theory and practice is within normal engineering accuracy.

Nomenclature

AT_c = ampere-turns to produce coercive force H_c
$AT_{c,dc}$ = ampere-turns to produce static hysteresis loop
E_C = control voltage, volts, average
E_0 = output voltage, volts, average
$E_{0,m}$ = maximum output voltage, volts, average
E_m = supply voltage, line-to-line voltage, maximum (crest)
$e_{C,A}$, $e_{C,C'}$, $\ldots e_{C,B'}$ = voltage across control winding of saturable reactors A, C', $\ldots B'$ respectively, volts, instantaneous
$e_{G,A}$, $e_{G,C'}$, $\ldots e_{G,B'}$ = voltage across gate winding of saturable reactors A, C', $\ldots B'$ respectively, volts, instantaneous
e_0 = output voltage, volts, instantaneous
e_{1-0}, e_{2-0}, e_{3-0} = line-to-neutral voltage, volts, instantaneous
e_{1-2}, e_{2-3}, e_{3-1} = line-to-line voltage, volts, instantaneous
$\hat{e}_0$ = per-unit output voltage, equation 37
f = supply frequency, cycles per second
H_c = dynamic coercive force, oersteds
I_C = control current, amperes, average
i_C = control current, amperes, instantaneous
i_L = load current, amperes, instantaneous
K_E = differential voltage gain, equation 16
m = factor for equation 22, also shown in Fig. 6
N = number of turns
N_C = number of turns of one control winding
N_G = number of turns of one gate winding
p = Heaviside operator
R_C = resistance of entire control circuit, ohms
R_G = resistance of gate winding, ohms
R_L = load resistance, ohms
R_0 = total resistance encountered by load current, ohms
R_f = rectifier forward resistance, ohms
T_C = time constant of control circuit, with no other circuits coupled, seconds
α = saturation (firing) angle, degrees or radians
ζ = boundary angle, equation 36b, degrees
θ = angular variable, degrees or radians
θ_d = dwell angle, degrees or radians
ϑ = average dwell, unity
ϕ_A, $\phi_{C'} \ldots \phi_{B'}$ = core flux in saturable reactors A, $C' \ldots B'$ respectively, maxwells, instantaneous
ψ_C = total flux linkage with control windings, maxwell turns, instantaneous
Ψ_B = total flux linkage with control windings, maxwell turns, average
ω = angular velocity of line voltage, radians per second

Incremental quantities are indicated by bar over symbol.

Appendix I

Conforming to equation 32A, the flux linkage change is

$$d\Psi_C = 6\left[(e_{G,A}\vartheta)\Big|^{60+\alpha} + (e_{G,A}\vartheta)\Big|^{120+\alpha} - (e_{G,A}\vartheta)\Big|^{180+\alpha}\right]\frac{N_C}{N_G}\frac{10^8}{\omega}\,d\alpha \qquad (38A)$$

From Figs. 4(C) and 1(B) and 1(C):

$e_{G,A}$	ϑ
$-\frac{2}{3}e_{3-1}\Big\|^{60+\alpha} = \frac{2}{3}e_{1-2}\Big\|^{\alpha}$ …………………………	$\frac{2}{3}$
$\left(e_{1-2}-\frac{1}{3}e_{2-3}\right)\Big\|^{120+\alpha} = -e_{1-2}\Big\|^{-60+\alpha} - \frac{e_{1-2}}{3}\Big\|^{\alpha}$ ……	$\frac{1}{2}$
$-\frac{1}{3}e_{1-2}\Big\|^{180+\alpha} = +\frac{1}{3}e_{1-2}\Big\|^{\alpha}$ …………………………	$\frac{1}{3}$

Substituting into equation 38A yields

$$\frac{d\Psi_C}{d\alpha} = 6\left(\frac{1}{6}e_{1-2}\Big|^{\alpha} - \frac{1}{2}e_{1-2}\Big|^{-60+\alpha}\right)\frac{N_C}{N_G}\frac{10^8}{\omega} \qquad (38B)$$

The change in the average of the output voltage dE_0

$$dE_0 = -\frac{3}{\pi}e_{1-2}\Big|^{-60+\alpha}d\alpha \qquad (38C)$$

$$\frac{dE_0}{d\alpha} = -\frac{3}{\pi}e_{1-2}\Big|^{-60+\alpha} \qquad (38D)$$

Substituting for e_{1-2} from equation 5A yields from the above equations and equation 27

$$\frac{d\Psi_C}{dE_0} = \left[1+\frac{1}{3}\,\frac{\sin(\alpha-60°)}{\sin(120°-\alpha)}\right]\frac{N_G}{N_G}\frac{10^8}{2f} \qquad (38E)$$

$$m = 1+\frac{1}{3}\,\frac{\sin(\alpha-60°)}{\sin(120°-\alpha)} \qquad (38F)$$

Appendix II

The change in flux linkage is, from equation 32A,

$$d\Psi_C = 6\left[(e_{G,A}\vartheta)\Big|^{120+\alpha} + (e_{G,A}\vartheta)\Big|^{180+\alpha} + (e_{G,A}\vartheta)\Big|^{240+\alpha}\right]\frac{N_C}{N_G}\frac{10^8}{\omega}\,d\alpha \qquad (39A)$$

from Figs. 4(D) and 1(B) and (C)

$e_{G,A}$	ϑ
$(2e_{2-3}-e_{3-1})\Big\|^{120+\alpha} = -2e_{3-1}\Big\|^{60+\alpha} - e_{3-1}\Big\|^{60-\alpha}$ ……	$\frac{1}{2}$
$e_{1-2}\Big\|^{180+\alpha} = e_{3-1}\Big\|^{60+\alpha}$ …………………………	$\frac{1}{3}$
$-e_{2-3}\Big\|^{240+\alpha} = e_{3-1}\Big\|^{60+\alpha}$ …………………………	$\frac{1}{6}$

Substituting into equation 39A yields

$$\frac{d\Psi_C}{d\alpha} = 6\left(-\frac{1}{2}e_{3-1}\Big|^{60+\alpha} - \frac{1}{2}e_{3-1}\Big|^{60-\alpha}\right)\frac{N_C}{N_G}\frac{10^8}{\omega} \qquad (39B)$$

The change of output voltage dE_0 consists of two components:

$$dE_0 = -\frac{3}{\pi}\left[e_{1-2}\Big|^{-60+\alpha} - e_{1-2}\Big|^{\alpha}\right]d\alpha \qquad (39C)$$

From Figs. 1(B) and (C) or equations 5A and 5C

$$e_{1-2}\Big|^{-60+\alpha} - e_{1-2}\Big|^{\alpha} = -e_{3-1}\Big|^{-60+\alpha} \qquad (39D)$$

$$dE_0 = -\frac{3}{\pi}(-e_{3-1})\Big|^{-60+\alpha}d\alpha \qquad (39E)$$

It will be seen that dE_0 is represented by the crosshatched area of Fig. 4(D).

Substituting for e_{3-1} from equation 5C, and substituting equations 39(B) and (E) into equation 27 yields

$$m = 1-\frac{\sin(60°-\alpha)}{\sin 60°} \qquad (39F)$$

Appendix III

The change in flux linkage is given by equation 39A. From Figs. 4(E) and 1(B) and (C)

$e_{G,A}$	ϑ
$-\frac{1}{3}e_{3-1}\Big\|^{120+\alpha} = \frac{1}{3}e_{3-1}\Big\|^{-60+\alpha}$ …………………………	$\frac{1}{2}$
$\frac{1}{3}e_{1-2}\Big\|^{180+\alpha} = \frac{1}{3}e_{3-1}\Big\|^{-60+\alpha}$ …………………………	$\frac{1}{3}$
$-\frac{1}{3}e_{2-3}\Big\|^{240+\alpha} = \frac{1}{3}e_{3-1}\Big\|^{-60+\alpha}$ …………………………	$\frac{1}{6}$

Substituting into equation 32B yields

$$\frac{d\Psi_C}{d\alpha} = 6\times\frac{1}{3}e_{3-1}\Big|^{-60+\alpha}\frac{N_C}{N_G}\frac{10^8}{\omega} \qquad (40A)$$

The change in dE_0 is again

$$dE_0 = -\frac{3}{\pi}(-e_{3-1})\Big|^{-60+\alpha}d\alpha \qquad (39E)$$

Substituting equations 40A and 39E into equation 27 yields

$$m = \frac{2}{3} \qquad (40B)$$

References

1. THEORY OF 3-PHASE BRIDGE MAGNETIC AMPLIFIER, I—STEADY STATE, H. F. Storm, C. W. Flairty. *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 78, 1959 (Jan. 1960 section), pp. 878–95.

2. THE TRANSDUCTOR AMPLIFIER (book), U. H. Krabbe. Lindhska Boktrycheriet, Orebro, Sweden, 1947.

3. THE TRANSDUCTOR (book), A. U. Lamm. Esselte Aktiebolaget, Stockholm, Sweden, second edition, 1948.

4. MAGNETIC AMPLIFIERS (book), H. F. Storm. John Wiley & Sons, Inc., New York, N. Y., 1955, p. 359.

5. MAGNETIC AMPLIFIERS (book, in Russian), M. A. Rosenblat. Isdatelstvo "Sovietskoe Radno," Moscow, USSR, 1956, p. 577.

6. MAGNETIC AMPLIFIERS (book, in Czechoslovakian), M. Gabler, J. Hańkovec, E. Tománek. Státní Nakladatelství Technické Literatury, Prague, Czechoslovakia, 1956, p. 359.

7. TRANSDUCTORS AND MAGNETIC AMPLIFIERS (book), A. G. Milnes. Macmillian and Company, Ltd., London, England, 1957, pp. 181–207.

8. DYNAMIC BEHAVIOR OF A 3-PHASE MAGNETIC AMPLIFIER, F. J. Ellert. *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 75, Nov. 1956, pp. 554–60.

9. THE POLYUNIT SATURABLE REACTOR, Karl I. Selin. *Ibid.*, pt. III (*Power Apparatus and Systems*), vol. 75, Oct. 1956, pp. 863–67.

10. EXPERIMENTAL CHARACTERISTICS OF THE 3-PHASE POLYUNIT SATURABLE REACTOR, Karl I. Selin, Alexander Kusko. *Ibid.*, pp. 868–71.

11. Magnetic Amplifiers for High-Performance Power Servos, G. P. DeWestfelt. *Electrical Manufacturing*, New York, N. Y., Nov. 1956, p. 114.

12. The Power Magnestat, S. Rudzinski. *B.T.H. Activities*, Rugby, England, vol. 27, no. 5, Sept.-Oct. 1956, pp. 154–58.

13. Three-Phase High-Speed Magnetic Amplifiers, A. E. Maine. *Electronic Engineering*, London, England, Dec. 1954, pp. 514–21.

14. 160,000-Ampere High-Speed Magnetic-Amplifier Design, A. B. Rosenstein. *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 74, Mar. 1955, pp. 90–99.

15. Generalized Transfer Characteristics for 3-Phase Magnetic Amplifiers with Inductive Loads, C. H. Grace. "*Proceedings*, Special Technical Conference on Magnetic Amplifiers," *AIEE Special Publication T98*, Aug. 1957, pp. 169–76.

16. New Techniques in Designing Three-Phase Magnetic Amplifiers, W. H. Moore. *Ibid.*, pp. 177–86.

17. Magnetic-Amplifier Design, Part I—The Core, R. E. Anderson. *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 77, May 1958, pp. 160–67.

18. Magnetic-Amplifier Design, Part II—The Windings, R. E. Anderson. *Ibid.*, pp. 167–75.

A Generalized Chemical Processing Model for the Investigation of Computer Control

THEODORE J. WILLIAMS
NONMEMBER AIEE

ROBERT E. OTTO
NONMEMBER AIEE

AT PRESENT there is no direct method of comparing the relative applicability of the many computer-control schemes which have recently been proposed for use by the chemical and petroleum processing industries. While many of the computer manufacturers involved have undertaken joint study programs with several of the chemical and petroleum companies, proprietary considerations have to date prevented a direct comparison of the results obtained.

For this reason the Monsanto Chemical Company has proposed the use of the chemical processing model described herein to those interested in computer control with the hope that it may serve as a basis for a direct comparison of the available computers and of their relative capabilities.

Every effort has been made to have this model include as many as possible of the effects present in typical chemical manufacturing plants. Reaction, heat exchange, multiple separation steps, and recycle streams have, therefore, been included, since these will usually be part of a typical process.

Monsanto will welcome all suggestions from any interested reader which may serve to increase the generality of the model. We will also appreciate suggestions for further uses to which this or other processing models may be put to increase the application and knowledge of advanced process-control principles and practices.

Statement of Problem

The plant considered here is to manufacture 40,000,000 pounds per year of a chemical product, *P*, whose exact nature need not be specified at this time. The chemical kinetics of the reaction involved in making *P* are well known and are described later along with a discussion of the reactor scheme to be used.

It will be assumed that a plant has already been designed and built for producing *P*. However, considerable difficulty has been experienced with the stability of operation of the plant, and it is desired to investigate the feasibility of applying computer control in an effort to achieve stable operation. At the same time, it is desired that the computer maintain an optimum balance of plant operating conditions to assure that the maximum return on investment is being obtained.

Existing instrumentation is indicated in the details of the various plant units given in the paper. All existing instruments are pneumatic. The process is equipped with a control house and all indicators and controllers are either mounted on the "flow-chart" type of panel board in the control house or have a remote reader at that location.

It has been requested that the computer manufacturers specify any changes in the the present instrumentation system which they feel are necessary. All new instrumentation specified must be capitalized with the computer. Any present instrumentation retained may be used without cost to the computer installation, since it has already been included in plant depreciation costs.

An extensive study has been made of the economics of this process and the results are indicated in the section on Cost Factors. It has been found that the plant, when at steady state, should operate at very close to the optimum return on investment. However, the instabilities encountered have prevented the attainment of such results, hence this request for a study of the applicability of computer control.

In the interests of reducing the possible work load in investigating this problem, it is requested that the computer manufacturers confine their efforts to the economics of plant operation. No effort should be spent on chemical engineering studies looking toward process improvement by the redesign of plant units or by the addition of further processing equipment. Since all companies considering this problem will not have equivalent chemical engineering staffs, inclusion of such studies, however desirable, would work a hardship on some participants in the study.

Plant Sizing

Fig. 1 shows a block diagram of the plant as now operated with the present flow rates and associated stream temperatures indicated. Each of the plant units is described in detail in the following sections of this report.

Table I presents the chemical analysis data and component flow rates for each stream in the plant.

Physical Properties

Appendix I and its associated Figs. 2–5 present the pertinent physical data on the

---

Paper 60-119, recommended by the AIEE Computing Devices Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., January 31–February 5, 1960. Manuscript submitted November 2, 1959; made available for printing April 7, 1960.

Theodore J. Williams and Robert E. Otto are with the Monsanto Chemical Company, St. Louis, Mo.

This model is a further development of the generalized model proposed by L. Cooper and T. J. Williams before the AIEE Workshop on Control Computers, University of Colorado, Boulder, Colo., May 21–22, 1959. The authors are indebted to D. P. Eckman of Case Institute of Technology for his many suggestions concerning the model and its use. The help of many of their associates at Monsanto is also gratefully acknowledged, especially that of William E. Ball.

Vol.
Pt. 1
196

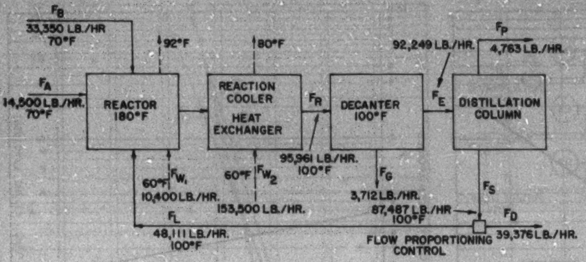


Fig. 1. Process showing relationship of process units, main stream rates, and stream temperatures

components and mixtures involved in this generalized process.

The vapor pressure of the reactor effluent mixture as given in Fig. 2 determines that the operating pressure of the reactor will be 3 atm (atmospheres) and that of the distillation column to be approximately 350 mm (millimeters) of mercury (Hg). It can be assumed that the vapor-pressure temperature curve does not vary greatly from that given in Fig. 2 as reactor conditions vary. This is because the fractional composition of components *A*, *B*, and *E* will always remain high under recycle conditions and all of these have properties very close to those of the example mixture. Therefore those properties ascribed to the reactor effluent mixture, F_R, may be used for any pertinent stream with negligible error.

Reactor and Chemical Reaction Involved

DETAILS OF CHEMICAL REACTION

P is produced by or takes part in the following chemical reactions:

1. $A+B \rightarrow C$, reaction coefficient, k_1 (1)
2. $C+B \rightarrow P+E$, reaction coefficient, k_2 (2)
3. $P+C \rightarrow G$, reaction coefficient, k_3 (3)

Initial reactant materials, *A* and *B*, are available in pure form from previous processing units of the over-all plant complex or outside sources as discussed later.

Components *C* and *E* are intermediates and/or by-products of the reaction. They have no sale value as chemical products but can be disposed of as fuels. The same is true of streams of mixed components such as the discard stream from the column bottoms.

By-product *G* is a heavy oily material which must be disposed of as a waste material.

The reaction coefficients k_1, k_2, and k_3 have been evaluated. Their values can be expressed in graphical form as in Fig. 6 or by the Arrhenius equation,

$$k_i = A_i e^{-B_i/T} \qquad (4)$$

where

$A_1 = 5.9755 \times 10^9$/hour, weight fraction
$B_1 = 12{,}000$ R (degrees Rankine), basis, 1 pound of *A* or *B*
$A_2 = 2.5962 \times 10^{12}$/hour, weight fraction
$B_2 = 15{,}000$ R, basis, 1 pound of *B*
$A_3 = 9.6283 \times 10^{15}$/hour, weight fraction
$B_3 = 20{,}000$ R, basis, 1 pound of *C*

Note that the volume dimension in the reaction coefficient, *k*, is "pound of mixture" rather than cubic feet or other strictly volume dimension. This is satisfactory since specific gravity is considered a constant in the reactor.

Heats of reaction of the three reactions are as follows:

Reaction 1: $H_1 = -125$ Btu per pound of *C* produced.

Reaction 2: $H_2 = -50$ Btu per pound of $E+P$ produced.

Reaction 3: $H_3 = -143$ Btu per pound of *G* produced.

That is, all reactions are exothermic.

DETAILS OF CHEMICAL-REACTOR DESIGN

Fig. 7 diagrams the pertinent details of the continuous stirred-tank reactor used for producing *P*. Appendix II lists the pertinent design data other than the kinetics already discussed.

Reaction-Cooler Heat Exchanger

To avoid an overproduction of the heavy oil waste material, *G*, and a subsequent reduction in product, *P*, it is necessary that the reaction be cooled to such a degree that the reaction is effectively stopped. In addition, as shown by Fig. 4, material *G* becomes insoluble in the reaction mixture when the temperature has been reduced below a certain value. Therefore, a cooler has been installed to reduce the temperature of the reaction mixture to 100 C (degrees centigrade) or below in order to satisfy both requirements. Fig. 8 diagrams the exchanger chosen and Appendix III gives its specifications.

Decanter

Since the heavy oil waste material, *G*, is insoluble at stream temperatures of 100 F (degrees Fahrenheit) or below, and since it has a considerably higher specific gravity than the carrier stream in which it is suspended, it may be removed by settling. The decanting vessel diagramed in Fig. 9 carries out this separation. The vessel is constructed of steel and need not be a pressure vessel as is the reactor. The effective liquid volume of 185 feet³ allows

Table I. Stream Rates and Composition in Various Locations in Generalized Process Model

	Stream																	
	F_A		F_B		F_R		F_E		F_G		F_P		F_S		F_L		F_D	
Component	Rate, Lb/Hr	Composition, Weight Fraction	Rate, Lb/Hr	Composition, Weight Fraction	Rate, Lb/Hr	Composition, Weight Fraction	Rate, Lb/Hr	Composition, Weight Fraction	Rate, Lb/Hr	Composition, Weight Fraction	Rate, Lb/Hr	Composition, Weight Fraction	Rate, Lb/Hr	Composition, Weight Fraction	Rate, Lb/Hr	Composition, Weight Fraction	Rate, Lb/Hr	Composition, Weight Fraction
A	14,500	1.000			11,628	0.121	11,628	0.126					11,628	0.133	6,393	0.133	5,235	0.133
B			33,350	1.000	36,829	0.383	36,829	0.399					36,829	0.421	20,255	0.421	16,574	0.421
C					2,348	0.025	2,348	0.026					2,348	0.027	1,289	0.027	1,059	0.027
E					33,348	0.346	33,348	0.361					33,348	0.381	18,340	0.381	15,007	0.381
P					8,096	0.085	8,096	0.088			4,763	1.000	3,334	0.038	1,834	0.038	1,501	0.038
G					3,712	0.039			3,712	1.000								
Total	14,500	1.000	33,350	1.000	95,961	1.000	92,249	1.000	3,712	1.000	4,763	1.000	87,487	1.000	48,111	1.000	39,376	1.000

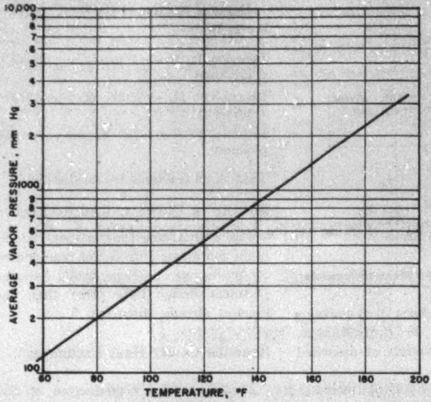


Fig. 2. Vapor pressure relationships for reactant mixture (less G) as feed to distillation column

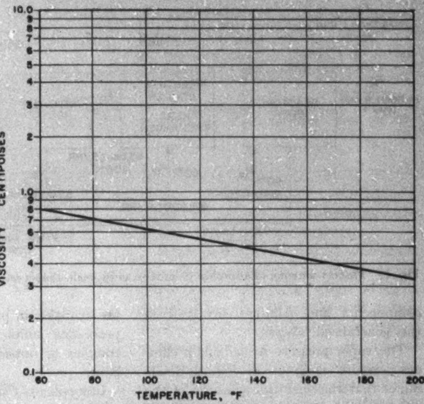


Fig. 3. Viscosity relationships for reactant mixture of composition equal to F_R stream

an average residence time of the liquid of 5.8 minutes which is sufficient to allow the required separation.

It must be kept in mind that should the temperature of the material in the decanter ever rise above that temperature at which G becomes insoluble, the interface between the column feed material and G will disappear and G will mix throughout the available volume. At that time, reactor effluent material will be discharged in the F_G stream and G will be included in the feed to the column and be returned to the reactor in the F_L stream. Thus, under improper temperature control, G could build up to high levels in all streams, except F_P.

Distillation Column

Product P is separated from the effluent stream from the decanter, F_R, by distillation. Specifications of the column are given in Appendix IV. A flow chart of pertinent column stream rates is shown in Fig. 10. Figs. 11 and 12 and Appendixes V and VI give the design data for column reboiler and condenser and Fig. 13 that for a typical tray. The distillation column has been designed in such a way that the column condenser (Fig. 11) is just limiting at the design production rate (Table I) for the maximum coolant temperature, 75 F. However, the column itself will handle up to 1.77 times the design capacity (column flood point) provided the condenser could handle the resulting heat duty. Therefore, since the coolant temperature varies quite radically both diurnally and seasonally (Figs. 24 and 25), the resulting plant capacity is likewise quite variable.

The required coolant temperature is a function of the vapor pressure existing at the condenser and the temperature at which product P condenses at that particular pressure. The base pressure (325 mm Hg and 100 F) is established by the need to suppress further reaction taking place outside the reactor. Condenser pressure is, therefore, the base pressure less the total pressure drop of the column and of the condenser itself combined. Graphs of the variations of these pressure drops with throughput are given in Figs. 14 and 15. The resulting vapor condensation temperatures can then be derived from Fig. 2. Fig. 16 summarizes these data by pre-

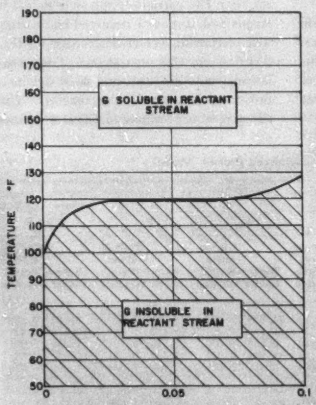
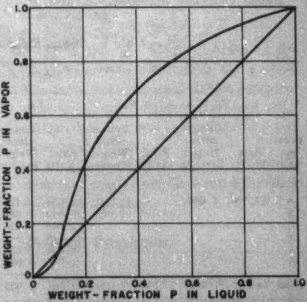
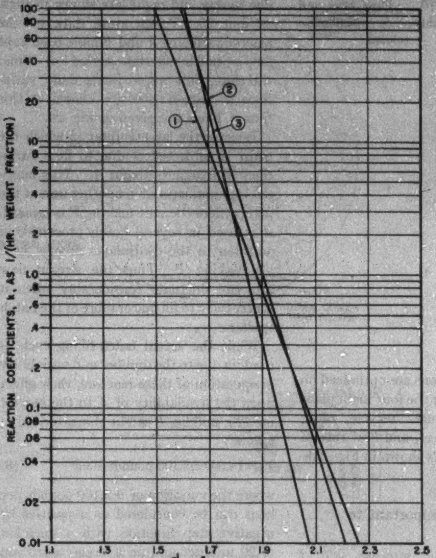


Fig. 4 (left). Solubility relationships for tar material, G, in plant streams

Fig. 5 (right). Vapor-liquid equilibrium relations for product material, P, and by-product material, E, as mixture





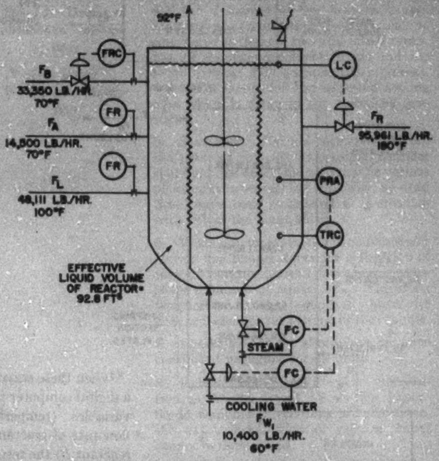

92°F
FRC
F_B
33,350 LB./HR.
70°F
FR
F_A
14,500 LB./HR.
70°F
FR
F_L
48,111 LB./HR.
100°F
LC
F_R
95,961 LB./HR.
180°F
PRA
TRC
EFFECTIVE LIQUID VOLUME OF REACTOR= 92.8 FT³
FC
STEAM
FC
COOLING WATER
F_{W_1}
10,400 LB./HR.
60°F

Fig. 6 (left). Reaction coefficients for reactions involved in generalized chemical reaction model for computer-control studies

Fig. 7 (above). Details of reaction vessel

senting the resulting required coolant temperature versus plant throughput as a function of the base design rate of Table I.

Fig. 17 diagrams the instrumentation used at present. As shown in Fig. 5, product P forms a high-boiling azeotrope with the intermediate by-product E. Thus, some P appears in the column bottoms stream, F_S, and therefore in the discard stream, F_D, and the recycle stream F_L.

Recycle Control

The instrumentation presently used to regulate the recycle and discard streams is illustrated in Fig. 18. Notice that the discard stream, F_D, is regulated as a certain fraction of the column bottoms flow, F_S. The recycle stream, F_L, then takes all excess and returns it to the reactor.

Cost Factors

This section itemizes the costs involved in the operation of the plant manufacturing P.

Appendix VII gives the raw material costs and conditions as well as selling prices and associated conditions. Appendix VIII lists the data used in deriving the manufacturing cost estimate, Fig. 19, while the over-all cost estimate for the plant operating case used is given in Table II.

Since the kinetics of this process are well known, it lends itself to extensive steady-state optimization. This has been done using the following equations:

$$\%\ \text{Return} = \frac{100\,[8400(M)-2.22(F_R)-0.124\,\{8400/(0.30F_P+0.0068F_D)\}-276{,}000]}{2{,}780{,}000} \qquad (5)$$

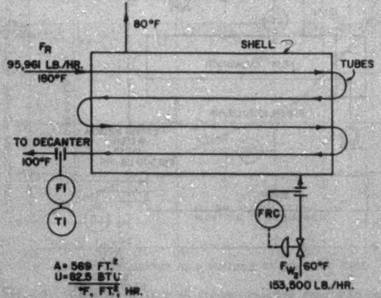

80°F
F_R
95,961 LB./HR.
180°F
SHELL
TUBES
TO DECANTER
100°F
FI
TI
FRC
F_{W_2}
60°F
153,500 LB./HR.
A= 569 FT.²
U=82.5 BTU / °F, FT.², HR.

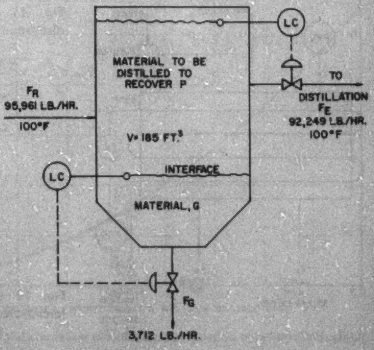
Fig. 8 (left). Details of reaction-cooler heat exchanger

Fig. 9 (right). Details of decanting vessel

Total volume = 185 ft³

Effective volume of column feed = 140 ft³ = V_R

Volume of G retained = 45 ft³ = V_E


LC
MATERIAL TO BE DISTILLED TO RECOVER P
F_R
95,961 LB./HR.
100°F
V= 185 FT.³
TO DISTILLATION
F_E
92,249 LB./HR.
100°F
LC
INTERFACE
MATERIAL, G
F_G
3,712 LB./HR.

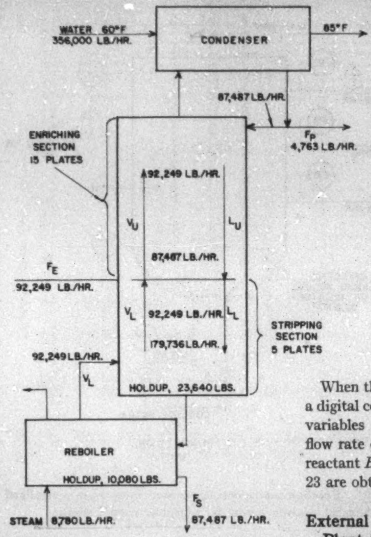

WATER 60°F
356,000 LB./HR.
CONDENSER
85°F
87,487 LB./HR.
F_P
4,763 LB./HR.
ENRICHING SECTION 15 PLATES
92,249 LB./HR.
V_U
L_U
87,487 LB./HR.
F_E
92,249 LB./HR.
V_L
92,249 LB./HR.
L_L
179,736 LB./HR.
STRIPPING SECTION 5 PLATES
92,249 LB./HR.
V_L
HOLDUP, 23,640 LBS.
REBOILER
HOLDUP, 10,080 LBS.
STEAM 8,780 LB./HR.
F_S
87,487 LB./HR.

Fig. 10. Flow rates and relationships in distillation column

where

$$M=0.30F_P+0.0068F_D-0.02F_A-0.03F_B-0.01F_G \quad (6)$$

and

$$F_P \leq 4{,}763 \text{ pounds/hour} \quad (7)$$

In these equations, M is the gross return per hour; \$276,000 represents the plant fixed charges, such as depreciation, labor, etc.; $2.22F_R$ represents utility charges which can be tied to plant flow rates; 0.124 (8400)$(0.30F_P+0.0068F_D)$ represent the SARE (sales, administration, research, and engineering) charges; and \$2,780,000 is the total investment involved.

When these equations are optimized on a digital computer for the four main plant variables (temperature, recycle ratio, flow rate of reactant A, and flow rate of reactant B) the results shown in Figs. 20–23 are obtained.

External Factors Important to Plant Operation

As mentioned earlier, this plant has shown considerable instability in its operation. A considerable amount of data has, therefore, been taken on those external variables which might have a bearing on its operation. Cooling water temperatures have been kept for some years at this location and the average annual variation of temperature due to the seasons and the average daily variation due to day and night, are shown in Figs. 24 and 25. Temperature data on the reactant streams A and B have been kept for a much shorter period of time; therefore, only approximate daily variations in temperature are known at this time. These are given in Fig. 26.

Perhaps the largest upset to which the plant is subjected is due to flow variations in reactant stream A. This material is manufactured in another unit of the plant where the unit making P is located. A is used in several batch processes, in addition to the continuous process used for making P. Thus the stream flow rate can fluctuate considerably when it is necessary to fill one or more of the batch reactors.

While the actual batch-filling cycle is random due to the number and variability of operation of these reactors, their effect upon the availability of A to the reactor used for making P can be characterized as follows:

$$F_A=14{,}500 \pm 2{,}100 \text{ pounds/hour} \quad (8)$$

where the variation of ±2,100 pounds per hour can be considered as a positive or negative step function with a duration of 20 minutes and an average repetition rate (one complete cycle, including both plus and minus variation) of once per hour.

Dynamic Equations

Because of the nature of the upsets imposed and the fact that plant operation has already been quite well optimized, the main need for computer control in

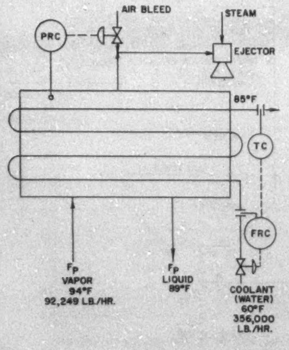

PRC
AIR BLEED
STEAM
EJECTOR
85°F
TC
FRC
F_P
VAPOR
94°F
92,249 LB./HR.
F_P
LIQUID
89°F
COOLANT
(WATER)
60°F
356,000
LB./HR.

Fig. 11 (left). Details of distillation column condenser

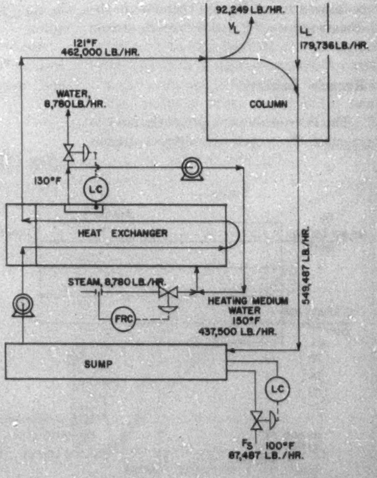

92,249 LB./HR.
V_L
L_L
179,736 LB./HR.
121°F
462,000 LB./HR.
COLUMN
WATER,
8,780 LB./HR.
130°F
LC
HEAT EXCHANGER
STEAM, 8,780 LB./HR.
FRC
HEATING MEDIUM
WATER
150°F
437,500 LB./HR.
549,487 LB./HR.
SUMP
LC
F_S 100°F
87,487 LB./HR.

Fig. 12 (right). Details of forced-circulation column reboiler

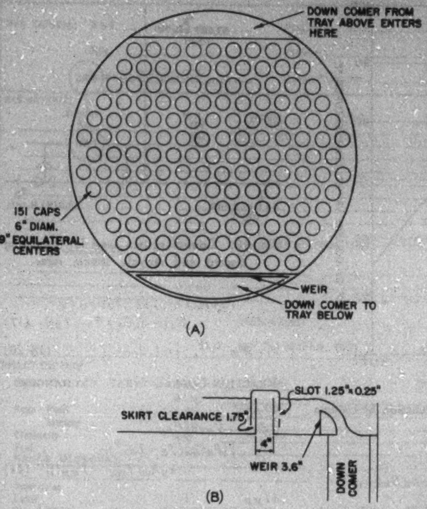


Fig. 13. Details of distillation column trays

A—Tray layout
B—Typical cross section

this plant is to smooth out the dynamic operation of the unit.

For this reason, a suggested set of differential equations is included here, which can be used to define the dynamic operation. The reader is, of course, free to use another set, either simpler or more complex, if he so desires.

SIMPLIFYING ASSUMPTIONS

Several simplifying assumptions may be made based upon knowledge of the plant design described earlier:

1. The reactor can be considered well stirred and reactants and products are at all times completely mixed.

2. The decanter can be considered well mixed in the layer of material serving as feed to the distillation column. There is no mixing between the light layer (V_E, distillation column feed) and the heavy layer (G) in the decanter.

3. Since P is readily separated from the other feed-stream components by the distillation column, the column may be treated as an instantaneous separator of F_P, i.e., F_E does not mix in the column. For purposes of computing column compositions as a function of time, the very approximate assumption may be made that the distillation column is a well-mixed tank with a volume equal to the total holdup of the column. Likewise, the reboiler may be considered as another well-mixed tank with a volume equal to its given holdup.

4. The dynamics of the heat exchanger for purposes of a first approximation may be considered as those of two well-mixed tanks, one for the reactor effluent stream and one for the coolant. This assumption requires a correction factor for the effective cooling area which is given in the equations presented.

5. Since every effort has been made to keep the plant as compact as possible, consider that fluid-line dead time between the various plant units is negligible. Likewise, no dead times have been considered in connection with any of the plant units.

6. Reboiler temperatures and consequently recycle line temperatures are a function of column operation and not of the temperature of the column feed stream, F_E. Therefore, temperature mixing equations have not been included for the column or the reboiler.

7. Loss of heat to the atmosphere has been considered negligible for all units.

8. Reboiler boilup has the same composition as reboiler liquid. Since the vapor-liquid equilibrium relations are not given, the exact composition of reboiler vapor cannot be determined.

The resulting equations are given in the following sections. All symbols are defined in the Nomenclature. Fig. 27 shows a block diagram of the actual dynamic system represented by the set of equations presented.

In these equations, all volumes will be expressed as pounds of contained liquid of a density of 50 pounds per cubic foot. All compositions will be expressed as weight fractions. Over-all reaction rates will be given as pounds of the component in question produced per hour per pound of reactor volume.

REACTOR EQUATIONS

$$\frac{dA_R}{dt}=\frac{F_A}{V_R}+\frac{F_L}{V_R}(A_L)-\frac{F_R}{V_R}(A_R)-k_1(A_R)(B_R) \tag{9}$$

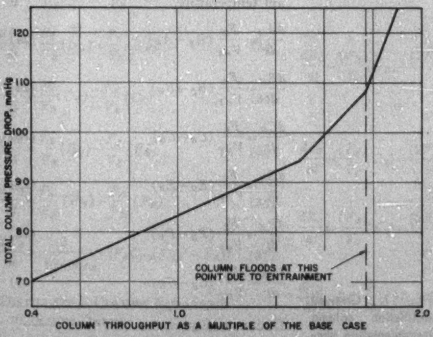


Fig. 14. Variation of over-all column pressure drop with throughput

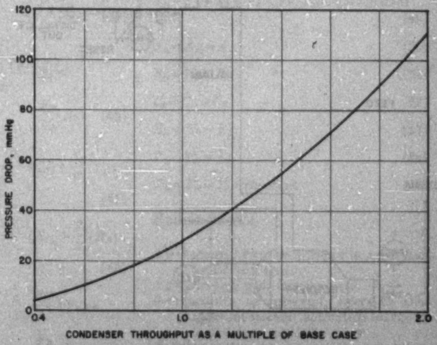


Fig. 15. Total condenser pressure drop as function of column throughput

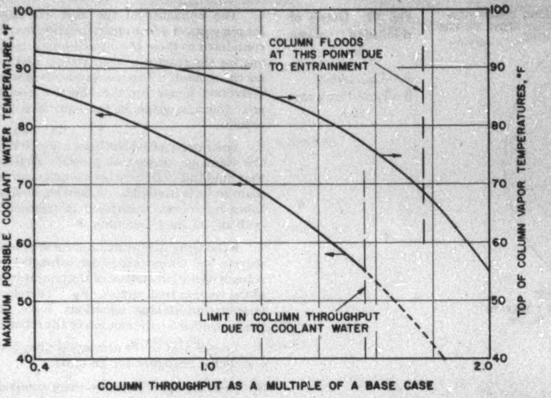


Fig. 16. Column vapor temperatures and cooling water requirements as function of column throughput

$$\frac{dB_R}{dt}=\frac{F_B}{V_R}+\frac{F_L}{V_R}(B_L)-\frac{F_R}{V_R}(B_R)-k_1(A_R)(B_R)-k_2(B_R)(C_R) \quad (10)$$

$$\frac{dC_R}{dt}=\frac{F_L}{V_R}(C_L)-\frac{F_R}{V_R}(C_R)+2k_1(A_R)(B_R)-2k_2(B_R)(C_R)-k_3(C_R)(P_R) \quad (11)$$

$$\frac{dE_R}{dt}=\frac{F_L}{V_R}(E_L)-\frac{F_R}{V_R}(E_R)+2k_2(B_R)(C_R) \quad (12)$$

$$\frac{dG_R}{dt}=1.5k_3(C_R)(P_R)-\frac{F_R}{V_R}(G_R) \quad (13)$$

$$\frac{dP_R}{dt}=\frac{F_L}{V_R}(P_L)-\frac{F_R}{V_R}(P_R)+k_2(B_R)(C_R)-0.5k_3(C_R)(P_R) \quad (14)$$

$$F_R=F_A+F_B+F_L \quad (15)$$

$$\text{Residence time per pass}=\frac{V_R}{F_R} \quad (16)$$

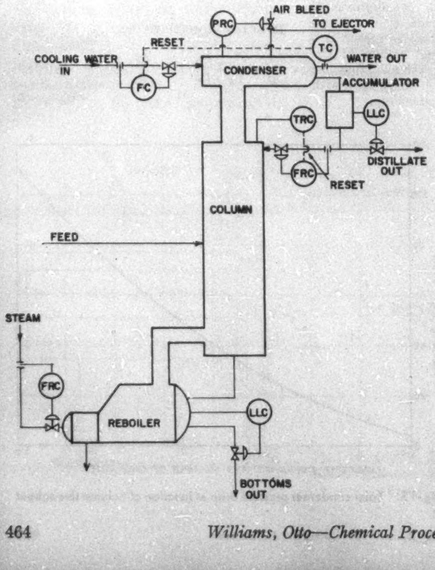

PRC
AIR BLEED
TO EJECTOR
RESET
TC
COOLING WATER IN
FC
CONDENSER
WATER OUT
ACCUMULATOR
TRC
LLC
FRC
RESET
DISTILLATE OUT
COLUMN
FEED
STEAM
FRC
REBOILER
LLC
BOTTOMS OUT

Fig. 17. Over-all distillation column instrumentation generalized process

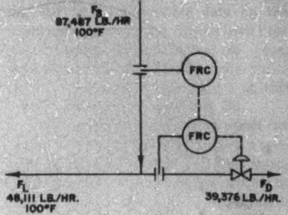

F_R
87,487 LB./HR
100°F
FRC
FRC
F_L
48,111 LB./HR.
100°F
F_D
39,376 LB./HR.

Fig. 18. Control system used for regulating discard and recycle stream rates

$$\frac{dT_R}{dt}=\frac{1}{V_R\rho_R C_{PR}}[2\{k_1(A_R)(B_R)\}V_R\{H_1\}+3\{k_2(B_R)(C_R)\}\{H_2\}V_R+1.5\{k_3(C_R)(P_R)\}\{H_3\}V_R-h_{10}A_{10}(T_R-T_W)+h_sA_s(T_s-T_R)-F_LC_{PL}(T_R-T_L)-F_AC_{PA}(T_R-T_A)-F_BC_{PB}(T_R-T_B)] \quad (17)$$

$$k_i=A_i\epsilon^{-B_i/T},\quad i=1,2,3 \quad (18\text{–}20)$$

REACTION-COOLER HEAT EXCHANGER

$$\frac{dT_{RH}'}{dt}=\frac{1}{V_{HT}\rho_R C_{PR}}\times\{[F_RC_{PR}(T_R-T_{RH}')]-h_H\bar{A}_H(T_{RH}'-T_{WH}')\} \quad (21)$$

$$\frac{dT_{WH}'}{dt}=\frac{1}{V_{HS}\rho_W C_{PW}}\times\{[F_WC_{PW}(T_W-T_{WH}')]+h_H\bar{A}_H(T_{RH}'-T_{WH}')\} \quad (22)$$

$$\bar{A}_H=\frac{A_H(T_{RH}'-T_{WH}')}{(T_{RH}'-T_{WH}')\ \text{ln mean}}=\text{effective area} \quad (23)$$

where

$$(T_{RH}'-T_{WH}')_{\text{ln mean}}=\frac{(T_R-T_{WH}')-(T_{RH}'-T_W)}{\ln\dfrac{(T_R-T_{WH}')}{(T_{RH}'-T_W)}}$$

and (T_R-T_{WH}') refers to the difference between entrance reactor effluent temperature and exit coolant temperatures, and $(T_{RH}'-T_W)$ refers to the difference between exit reactant stream temperature from the heat exchange and entrance coolant temperature.

$$\frac{dA_H}{dt}=\frac{F_R}{V_{HT}}(A_R-A_H) \quad (24)$$

$$\frac{dB_H}{dt}=\frac{F_R}{V_{HT}}(B_R-B_H) \quad (25)$$

$$\frac{dC_H}{dt}=\frac{F_R}{V_{HT}}(C_R-C_H) \quad (26)$$

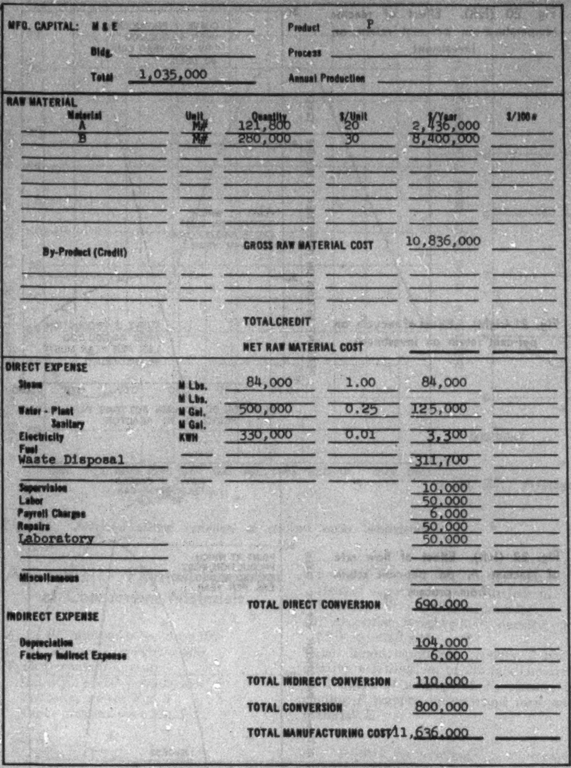
$$\frac{dE_H}{dt}=\frac{F_R}{V_{HT}}(E_R-E_H) \quad (27)$$

$$\frac{dP_H}{dt}=\frac{F_R}{V_{HT}}(P_R-P_H) \quad (28)$$

$$\frac{dG_H}{dt}=\frac{F_R}{V_{HT}}(G_R-G_H) \quad (29)$$

DECANTER EQUATIONS

$$F_G=G_RF_R \quad (30)$$



MFG. CAPITAL: M & E ______ Product P
Bldg. ______ Process ______
Total 1,035,000 Annual Production ______

RAW MATERIAL Material	Unit	Quantity	$/Unit	$/Year	$/100#
A	M#	121,800	20	2,436,000	
B	M#	280,000	30	8,400,000	
		GROSS RAW MATERIAL COST		10,836,000	
By-Product (Credit)					
		TOTALCREDIT			
		NET RAW MATERIAL COST			
DIRECT EXPENSE					
Steam	M Lbs.	84,000	1.00	84,000	
	M Lbs.				
Water - Plant	M Gal.	500,000	0.25	125,000	
Sanitary	M Gal.				
Electricity	KWH	330,000	0.01	3,300	
Fuel					
Waste Disposal				311,700	
Supervision				10,000	
Labor				50,000	
Payroll Charges				6,000	
Repairs				50,000	
Laboratory				50,000	
Miscellaneous					
		TOTAL DIRECT CONVERSION		690,000	
INDIRECT EXPENSE					
Depreciation				104,000	
Factory Indirect Expense				6,000	
		TOTAL INDIRECT CONVERSION		110,000	
		TOTAL CONVERSION		800,000	
		TOTAL MANUFACTURING COST		11,636,000	

Fig. 19. Manufacturing cost estimate

$$F_E = F_R - F_G \qquad (31)$$

$$\frac{dA_E}{dt} = \frac{F_R}{V_E}(A_H) - \frac{F_E}{V_E}(A_E) \qquad (32)$$

$$\frac{dB_E}{dt} = \frac{F_R}{V_E}(B_H) - \frac{F_E}{V_E}(B_E) \qquad (33)$$

$$\frac{dC_E}{dt} = \frac{F_R}{V_E}(C_R) - \frac{F_E}{V_E}(C_E) \qquad (34)$$

$$\frac{dE_E}{dt} = \frac{F_R}{V_E}(E_H) - \frac{F_E}{V_E}(E_E) \qquad (35)$$

$$\frac{dP_E}{dt} = \frac{F_R}{V_E}(P_H) - \frac{F_E}{V_E}(P_E) \qquad (36)$$

$$\frac{dT_E}{dt} = \frac{1}{V_E \rho_E C_{PE}}[F_E C_{PH}(T_{RH}' - T_E)] \qquad (37)$$

DISTILLATION COLUMN EQUATIONS (COLUMN ONLY)

$$F_P = (P_E - 0.1E_E)F_E \qquad (38)$$

$$\frac{dA_C}{dt} = \frac{(F_E)}{V_C}(A_E) - \frac{(2F_E - F_P)}{V_C}(A_C) + \frac{F_E}{V_C}(A_S) \qquad (39)$$

$$\frac{dB_C}{dt} = \frac{(F_E)}{V_C}(B_E) - \frac{(2F_E - F_P)}{V_C}(B_C) + \frac{F_E}{V_C}(B_S) \qquad (40)$$

$$\frac{dC_C}{dt} = \frac{(F_E)}{V_C}(C_E) - \frac{(2F_E - F_P)}{V_C}(C_C) + \frac{F_E}{V_C}(C_S) \qquad (41)$$

$$\frac{dE_C}{dt} = \frac{(F_E)}{V_C}(E_E) - \frac{(2F_E - F_P)}{V_C}(E_C) + \frac{F_E}{V_C}(E_S) \qquad (42)$$

$$\frac{dP_C}{dt} = 0.1\frac{(F_E)}{V_C}(E_E) - \frac{(2F_E - F_P)}{V_C}(P_C) + \frac{F_E}{V_C}(P_S) \qquad (43)$$

Table II. Analog Potentiometer Settings

Potentiometer No.	Value	Setting
1	F_A	0.1000
2		0.7500
3	P_S/E_S	0.1000
4		0.2500
5	F_B/F_A	variable
6	F_A	0.1000
7	P_S/E_S	0.1000
8	T_R-580	variable
9		0.2000
10		0.5800
11		0.4000
12, 13, 14	$50 \log (0.2V_R/F_A)$	variable
15	$50\ B_1 \log \epsilon$	0.1303
16	$50\ B_2 \log \epsilon$	0.1629
17	$50\ B_3 \log \epsilon$	0.2172
18	$50 \log A_1$	0.3233
19	$50 \log A_2$	0.3073
20	$50 \log A_3$	0.6941
21, 22, 23		0.0100
24	F_L/F_S	variable
25	P_S/E_S	0.1000
26		0.5000
27		0.0222
28		0.0084
29		0.5400
30		0.1680
31		0.2208

DISTILLATION COLUMN EQUATIONS (REBOILER ONLY)

$$\frac{dA_S}{dt} = \frac{(2F_E - F_P)}{V_S}(A_C) - \frac{F_S}{V_S}(A_S) - \frac{F_E}{V_S}(A_S) \qquad (44)$$

$$\frac{dB_S}{dt} = \frac{(2F_E - F_P)}{V_S}(B_C) - \frac{F_S}{V_S}(B_S) - \frac{F_E}{V_S}(B_S) \qquad (45)$$

$$\frac{dC_S}{dt} = \frac{(2F_E - F_P)}{V_S}(C_C) - \frac{F_S}{V_S}(C_S) - \frac{F_E}{V_S}(C_S) \qquad (46)$$

$$\frac{dE_S}{dt} = \frac{(2F_E - F_P)}{V_S}(E_C) - \frac{F_S}{V_S}(E_S) - \frac{F_E}{V_S}(E_S) \qquad (47)$$

$$\frac{dP_S}{dt} = \frac{(2F_E - F_P)}{V_S}(P_C) - \frac{F_S}{V_s}(P_S) - \frac{F_E}{V_S}(P_S) \qquad (48)$$

RECYCLE AND DISCARD LINES

$$F_S = F_D + F_L = F_E - F_P,\ F_D/F_L = \text{constant} \qquad (49)$$

$$A_S = A_D = A_L \qquad (50)$$

$$B_S = B_D = B_L \qquad (51)$$

$$C_s = C_D = C_L \qquad (52)$$

$$E_S = E_D = E_L \qquad (53)$$

$$P_S = P_D = P_L \qquad (54)$$

$$T_L = T_D = T_s = \text{constant} \qquad (55)$$

Nomenclature

A = reactant; frequency factor in Arrhenius equation; area
$\bar{A}_H$ = effective heat-transfer area
B = reactant; as subscript, refers to this reactant stream; activation energy in Arrhenius equation
C = intermediate component; as subscript, refers to distillation column

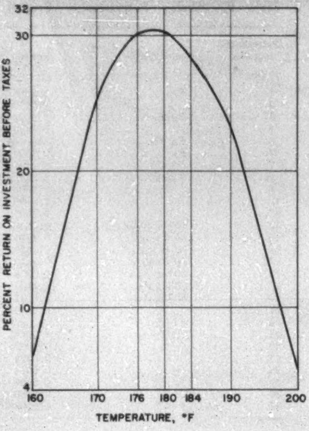
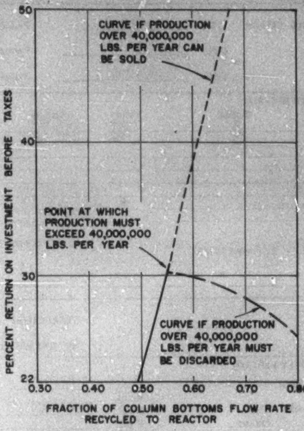


Fig. 20 (left). Effect of reactor temperature on per-cent return on investment

Fig. 21 (right). Effect of recycle on per-cent return on investment



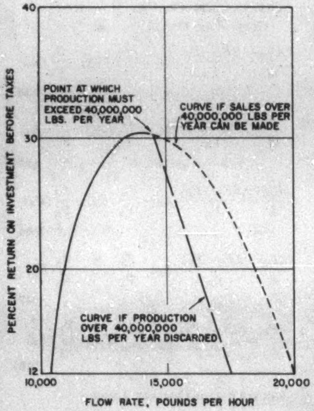
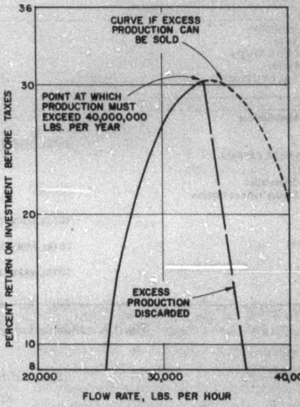


Fig. 22 (left). Effect of flow rate of reactant A on per-cent return from process

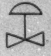
Fig. 23 (right). Effect of flow rate of reactant B on return on investment

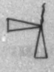


C_P = heat capacity; subscript gives stream
D = as subscript, refers to discard stream
E = intermediate component; as subscript, refers to decanter material used as distillation column feed
F = flow rate; subscript refers to origin of stream or its identity
G = by-product, heavy oil; as subscript, refers to heavy oil stream
H = heat of reaction; subscript gives reaction; as subscript, refers to reaction-cooler heat exchanger; second subscript identifies material involved
L = as subscript, refers to recycle stream
M = thousand; net hourly value of sales less raw material and waste-disposal charges
P = product; as subscript, refers to product stream
R = thermodynamic factor; as subscript, refers to reactor
S = as subscript, refers to column reboiler
T = temperature, may be F or R, depending on application; subscript defines location or stream involved
T' = average reaction-cooler heat-exchanger temperatures; subscript defines stream and location involved
V = volume; subscript defines location
W = as subscript, refers to water used as coolant
h = over-all heat-transfer coefficient
k = reaction rate constant; subscript defines reaction step considered
s = as subscript, refers to steam
t = time, usually in hours
RH = reaction effluent mixture in tube side of reaction-cooler heat exchanger
WH = coolant in shell side of reaction-cooler heat exchanger
HT = tube side of reaction-cooler heat exchanger
HS = shell side of reaction-cooler heat exchanger
ln mean = logarithmic mean (natural logarithms)
ϵ = base of Naperian logarithm
ρ = specific gravity referred to reactor effluent mixture of density equal to 50 pounds per cubic foot

The following are from the illustrations:

FR = flow recorder
FRC = flow recorder-controller
FC = flow controller
LC or LLC = level controller
PRA = pressure-recorder alarm
TRC = temperature-recorder controller
PRC = pressure-recorder controller

 = valve

 = pressure-relief valve

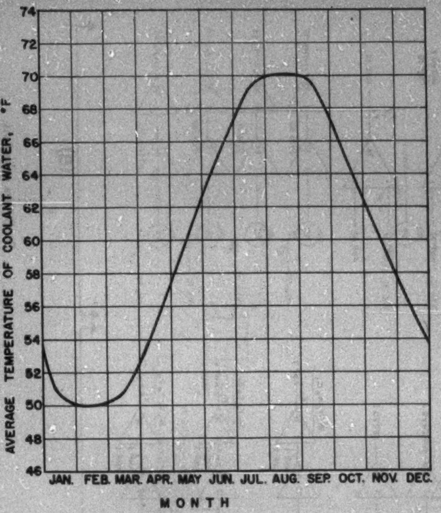


Fig. 24. Average annual variation of coolant water temperature

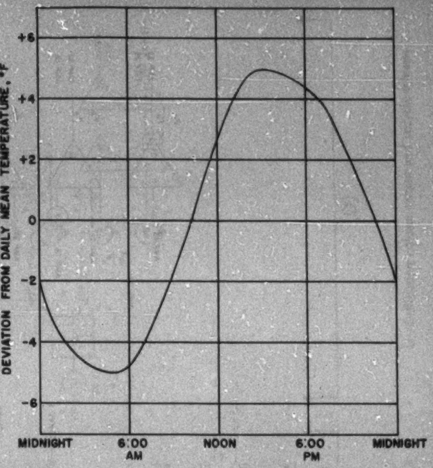


Fig. 25. Average diurnal variation of coolant water temperature

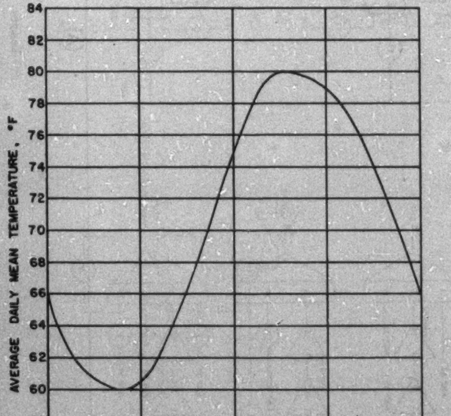
Appendix I. Physical Properties of Component Materials

A, B, P: molecular weight = 100
C, E: molecular weight = 200
Typical reactant mixture:
Density, $\rho = 50$ pounds/foot^3
Viscosity, μ = see Fig 3
Vapor pressure = see Fig. 2
$k = 0.086$ Btu/(hour)(feet^2)(F/foot)
Heat of vaporization = 95 Btu/pound
Heat capacity = 0.40 Btu/pound, F
G: molecular weight = 300; density, ρ = 68.6 pounds/foot^3

Vapor liquid equilibria: product P has a relative volatility, α, of 2.2 in comparison with liquid of average of F_S stream, including P trapped in azeotrope form with material E.

Appendix II. Design Data from Reactor

Effective liquid volume = 4,640 pounds
Excess heat produced/hour:
Gross materials formed: 18,540 pounds C/hour; 15,007 pounds E/hour; 7,501 pounds P/hour; 3,721 pounds G/hour.
Heat of reaction involved: 3,973,616 Btu/hr.
Heat required to bring input streams to reaction temperature: 3,644,952 Btu/hour.
Excess heat produced: 328,564 Btu/hour
Cooling coils installed:
382 foot BWG (Birmingham wire gage) *18*, 1-inch copper
$U = 50$ Btu/foot^2, F, hour
Effective coolant volume = 1.70 feet^3
Effective area = 100 feet^2



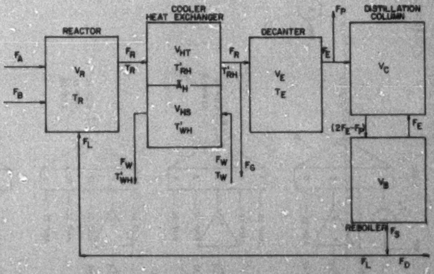

REACTOR
COOLER HEAT EXCHANGER
DECANTER
DISTILLATION COLUMN
REBOILER
F_A
F_B
V_R
T_R
F_R
T_R
F_L
V_{HT}
T'_{RH}
T_H
V_{HS}
T'_{WH}
F_W
T'_{WH}
F_R
T'_{RH}
F_W
T_W
F_G
V_E
T_E
F_E
F_P
V_C
$(2F_E - F_P)$
F_E
V_B
F_S
F_L
F_D

Fig. 26 (left). Average diurnal variation of temperature of reactant streams A and B

Fig. 27 (above). Approximate dynamic representation of chemical plant

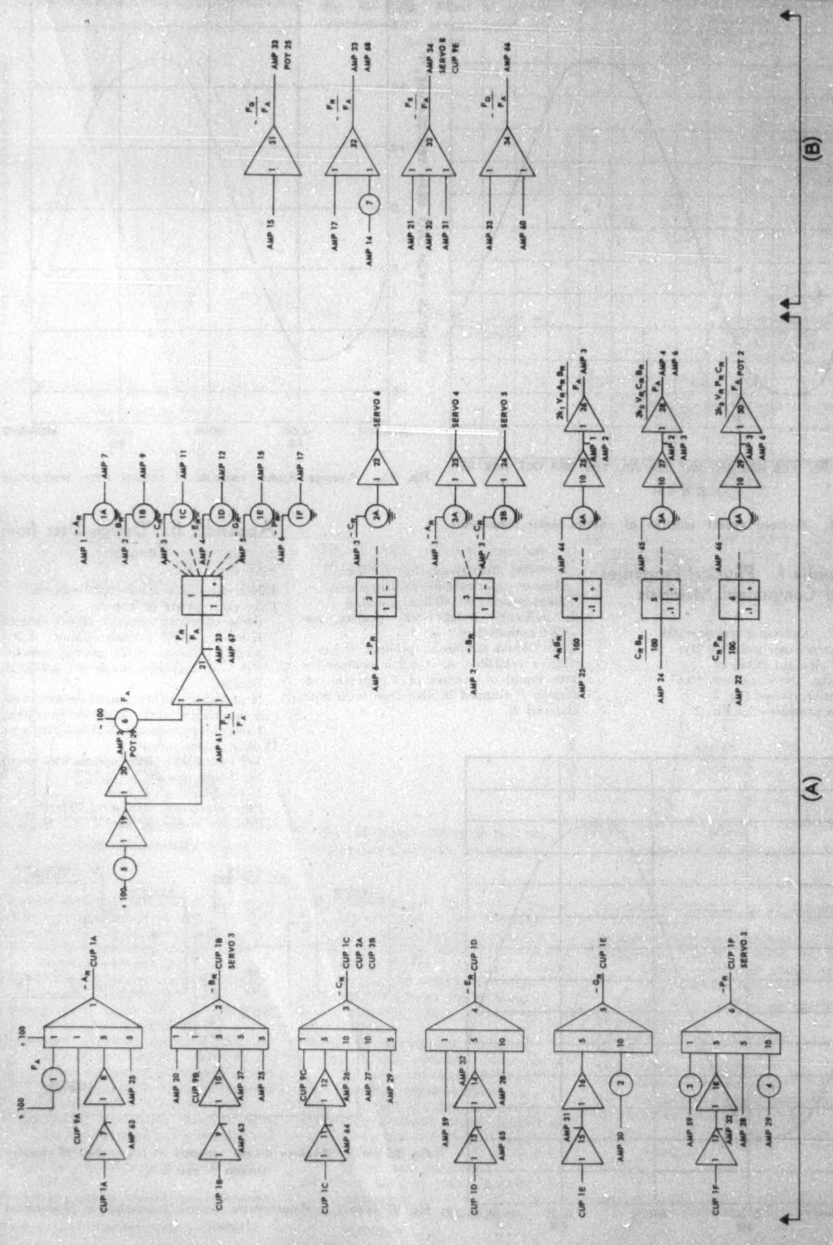

(A)
(B)
F_A
$-A_R$
$-B_R$
$-C_R$
$-E_R$
$-G_R$
$-P_R$
CUP 1A
CUP 1B
CUP 1C
CUP 1D
CUP 1E
CUP 1F
SERVO 3
SERVO 4
SERVO 5
SERVO 6
SERVO 8
F_R/F_A
F_L/F_A
$-F_G/F_A$
$-F_R/F_A$
$-F_S/F_A$
$-F_D/F_A$
$A_R B_R/100$
$C_R B_R/100$
$C_R P_R/100$
$2k_1 V_R A_R B_R/F_A$
$2k_2 V_R C_R B_R/F_A$
$2k_3 V_R P_R C_R/F_A$
+100
-100
POT 2
POT 25
AMP 1
AMP 2
AMP 3
AMP 15
AMP 17
AMP 21
AMP 33
AMP 60

Fig. 28. Analog wiring diagram

A—Reactor

B—Stream equations

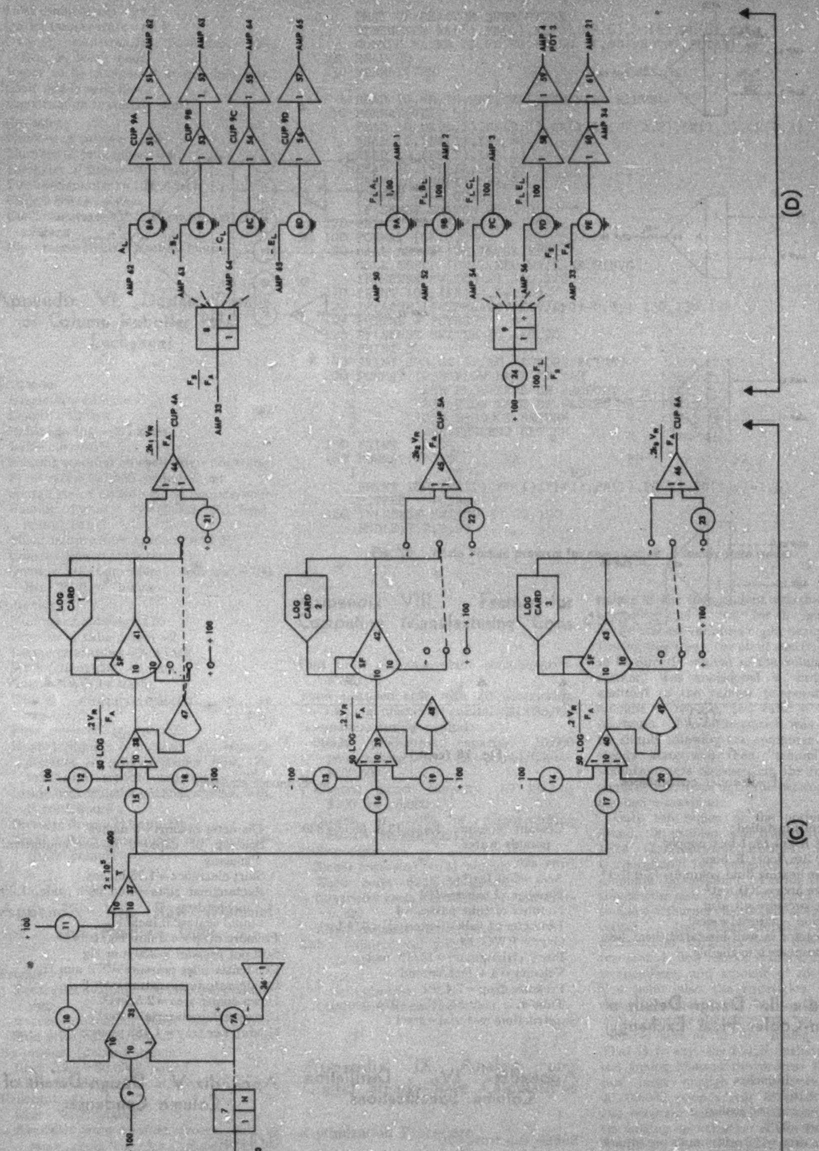

(C)
(D)
LOG CARD 1
LOG CARD 2
LOG CARD 3
50 LOG .2 V_R / F_A
2 × 10^5 / T − 400
T − 880
.2k1 V_R / F_A
CUP 4A
.2k2 V_R / F_A
CUP 5A
.2k3 V_R / F_A
CUP 6A
F_S / F_A
AMP 33
100 F_L / F_S
CUP 9A
CUP 9B
CUP 9C
CUP 9D
AMP 62
AMP 63
AMP 64
AMP 65
AMP 42
AMP 43
AMP 44
AMP 45
A_L
−B_L
−C_L
−E_L
F_L A_L / 100
F_L B_L / 100
F_L C_L / 100
F_L E_L / 100
AMP 1
AMP 2
AMP 3
AMP 4
POT 3
AMP 21
AMP 34
AMP 50
AMP 52
AMP 54
AMP 56
+100
−100
±100

Fig. 28 (cont.)

C—Reaction rate coefficients

D—Distillation column output and recycle streams

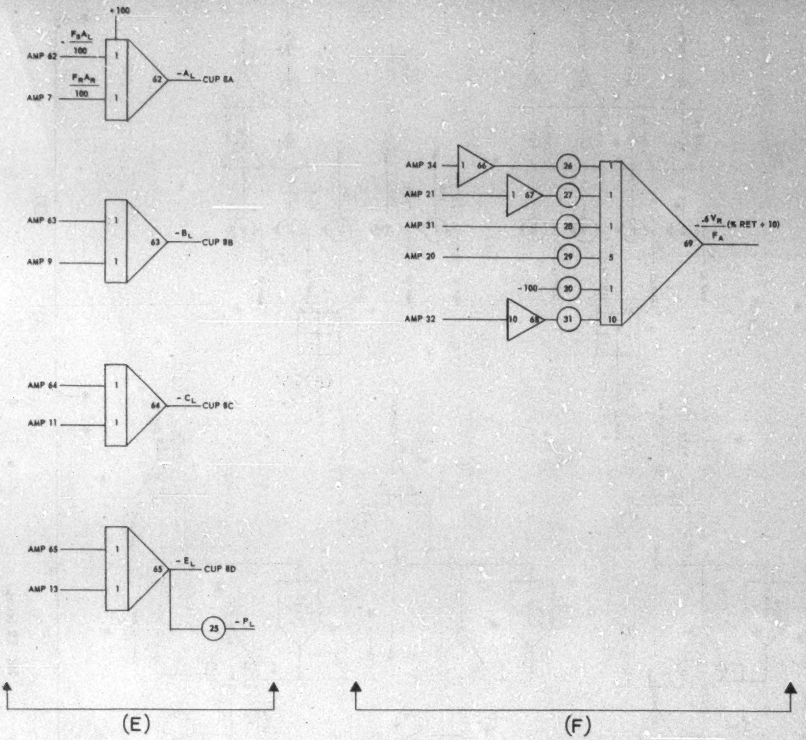

+100
$-\frac{F_BA_L}{100}$
AMP 62
1
$\frac{F_RA_R}{100}$
AMP 7
1
62
$-A_L$
CUP 8A
AMP 63
1
AMP 9
1
63
$-B_L$
CUP 8B
AMP 64
1
AMP 11
1
64
$-C_L$
CUP 8C
AMP 65
1
AMP 13
1
65
$-E_L$
CUP 8D
25
$-P_L$
(E)
AMP 34
1
66
26
AMP 21
1
67
27
AMP 31
28
AMP 20
29
-100
30
AMP 32
10
68
31
1
1
1
5
1
10
69
$-\frac{.6V_R}{F_A}$ (% RET + 10)
(F)

Fig. 28 (cont.)

E—Reboiler compositions

F—Economic equation

Heating coils installed:
- 382 feet BWG *18*, 1-inch copper
- $U = 100$ Btu/foot^2, F, hour
- Effective heating fluid volume = 1.70 feet^3
- Effective area = 100 feet^2

Operating pressure = 3 atm
Construction material = steel

Since reactor is well insulated, heat loss to the atmosphere is negligible.

Appendix III. Design Details of Reaction-Cooler Heat Exchanger

Shell side:
- Diameter = 16 inches
- Length = 16 feet
- Baffle spacing = 16 inches
- Baffle cut = 35%
- Pressure drop = 2.2 psi (pounds per square inch)
- Flow rate = 153,500 pounds/hour, water
- Inlet temperature = 60 F
- Outlet temperature = 80 F
- Over-all heat-transfer coefficient = 82.5 Btu, F, feet^2, hour
- Coolant volume, $V_{HS} = 13.3$ feet^3 = 830 pounds water

Tube side:
- Area = 569 feet^2
- Number of tubes = 184
- Number of tube passes = 4
- Diameter of tubes (external) = 3/4 inch
- Gage = BWG *14*
- Pitch (triangular) = 15/16 inch
- Velocity = 4.4 feet/second
- Pressure drop = 7.4 psi
- Tube-side volume, $V_{HT} = 274$ pounds

Construction material = steel

Appendix IV. Distillation Column Specifications

Bubble cap trays = 20
Plate spacing = 2 feet
Vapor velocity = 4.5 feet/second
Plate diameter = 10.0 feet
Weir length = 6 feet
Weir head = 0.136 foot or 1.65 inches
Bubble caps = 151 per tray
- Diameter of caps = 6 inches
- Spacing of caps = 9 inches equilateral spacing
- Skirt clearance = 1.75 inches
- Rectangular slots = 0.25 inch wide, 1.25 inches deep
- Liquid seal = 2.0 inches

Pressure drop = 4.0 mm Hg/tray
Still pot pressure = 350 mm Hg
Condenser inlet pressure = 270 mm Hg
Condensate temperature = 89 F
Down-comer area = 2.5 feet^2
Construction material = steel
Holdup per tray = 1,182 pounds

Appendix V. Design Details of Column Condenser

Shell side:
- Diameter = 39 inches
- Length = 20 feet
- Baffle spacing = 3 feet
- Baffle cut = 45%
- Pressure drop = 28 mm Hg
- Flow rate = 92,249 pounds/hour

Inlet temperature = 94 F
Outlet temperature = 89 F
Over-all heat-transfer coefficient = 150 Btu, F, feet², hour
Vapor to be condensed is on shell side
Shell side is split flow
Construction material = steel

Tube side:
Number of tubes = 1,258
Number of tube passes = 4
Diameter of tubes = 3/4 inch
Pitch (triangular) = 15/16 inch
Gage = BWG *14* steel
Coolant = water at 75 F maximum, 60 F average
Flow rate = 356,000 pounds/hour

Appendix VI. Design Details of Column Reboiler Heat Exchanger

Shell side:
Diameter = 40 inches
Length = 12 feet
Baffle spacing = 20 inches
Baffle cut = 25%
Heating medium on shell side = hot water
Flow rate = 437,500 pounds/hour
Sparge steam added = 8,780 pounds/hour
Fouling factor, R_d = (0.003 F, feet², hours)/Btu
Mean temperature gradient = 31 F
Construction material = steel
Outside heat-transfer coefficient = 700 Btu, F, feet², hours

Tube side:
Number of tubes = 1,176
Number of tube passes = 2
Diameter of tubes = 3/4 inch
Pitch (triangular) = 1 inch
Gage = BWG *14* steel
Heated medium = material to be vaporized
Flow rate = 462,000 pounds
Heated liquid is flashed at reduced pressure to provide vapor flow, V_L, to column; amount vaporized = 20%
Inside heat-transfer coefficient = 260 Btu, F, feet², hours
Pressure drop = 111 mm Hg
Total holdup (exchanger plus sump) = 10,080 pounds

Appendix VII. Raw Material and Product Data

Product, *P:*
Total sales available = 40,000,000 pounds/year
Excess production must be discarded
Sale price = $0.30/pound

By-product (discard stream):
Fuel value = $0.0068/pound
All available may be utilized

Reactant, *A:*
Cost = $0.02/pound
Available from another process unit in same plant complex. Supply under some restrictions as shown in equation 8

Reactant, *B:*
Cost = $0.03/pound
Available from external supplier in whatever quantity desired

```
C     TEST OF REACTOR SUBROUTINE
      DIMENSION FA(6),FB(6),FL(6),FR(6),FS(6),FD(6),CK(2,3)
      COMMON FA,FB,FL,FR,FS,FD,SG,SP,SD,PRODRT,PCTRET
  190 READ 30
   30 FORMAT(72H
     1                )
    1 READ 10,VR,TR,EPS,CMPL,ZED,SARE,RINVS
   10 FORMAT(7E10.5)
      READ 10,CK(1,1),CK(2,1),CK(1,2),CK(2,2),CK(1,3),CK(2,3)
      READ 10,(FA(I),I=1,6)
      READ 10,(FB(I),I=1,6)
      READ 10,(FL(I),I=1,6)
      READ 10,(FR(I),I=1,6)
      PCTPRE = -1.0E+35
      IF(SENSE SWITCH 2) 170,20
  170 PRINT 100, (FR(I),I=1,6)
  100 FORMAT (6E15,5)
   20 CALL REACTR(VR,TR,CK,EPS)
      CALL SSSEP (CMPL,ZED,VR,SARE,RINVS)
      IF(SENSE SWITCH 2) 110,120
  110 PRINT 100,(FR(I),I=1,6)
  120 IF(ABSF((PCTPRE-PCTRET)/EPS)-0.25) 130,130,124
  124 PCTPRE = PCTRET
  125 IF(SENSE SWITCH 3) 130,20
  130 PRINT 30
  140 PRINT 200,SG,SP,SD, PRODRT,PCTRET
  200 FORMAT (23HOFLOW RATE OF TAR     =   1PE15.5/
     1        23H FLOW RATE OF PRODUCT =   1PE15.5/
     2        23H FLOW RATE OF WASTE   =   1PE15.5/
     3        23H PRODUCT RETURN       =   1PE15.5/
     4        23H PERCENT RETURN       =   1PE15.5)
  150 PRINT   160
  160 FORMAT(120HO      FA             FB             FL            FR
     1            FS             FD                                       )
      PRINT 100,(FA(I),FB(I),FL(I),FR(I),FS(I),FD(I),I=1,6)
      PCTPRE = PCTRET
  180 IF(SENSE SWITCH 3) 20,190
      END(2,2,2,2,2)
```

Fig. 29. Main Fortran program for computation of steady-state values

Appendix VIII. Factors for Computing Manufacturing Costs

Unit price of main plant components = $382,000
Total installed cost with all accessories, i.e., manufacturing capital = $1,035,000

Nonmanufacturing capital:
Steam production capacity, 10,000 pounds/hour at $20/pound/hour = $200,000
Electric power capacity, 50 kw at $100/kw = 5,000
Others, at 1.0% of manufacturing costs = 116,000; $321,000

Utility costs:
Steam costs = $1.00/M pounds
Water costs = $0.25/M gallons
Electricity costs = $0.01/kw-hr (kilowatt-hour)
Waste disposal = $0.01/pound *G*

Labor required, per shift:
1 operator
1 helper
1 plant chemist
1 part-time supervisor

Appendix IX. Analog and Digital Steady-State Calculations

Optimization Procedure

In the steady-state optimization of the process under discussion, the analog and digital computers were used sequentially. Initially, the steady-state process equations and the economic equations were solved on the analog computer in order that wide ranges of the independent variables would be studied and a "feel" for the importance of the process variables could be developed. During this stage, the plant size (considered to be directly related to the volume of the reactor) was considered a variable, in addition to the reactor temperature, the amount of recycle, and ratio of the two reactants. The optimization was achieved by simply allowing the computer to reach steady state and then "trimming" the potentiometers representing the four independent variables, until the per-cent return reached a maximum.

Once the region of the optimum was found, the problem was transferred to the IBM (International Business Machines Corporation) *704* computer in order to pinpoint the optimum. During this stage, the process size (reactor volume) was fixed to allow a sharper calculation of the process economics and design details. The four independent variables were flow rate of reactant *A*, flow rate of reactant *B*, reactor temperature, and amount of recycle. At this point also the restriction that the excess product must be discarded was added because the plant was no longer of variable size as in the analog program. That is to say, the rough optimization on the analog allowed the process size to be just large enough to produce exactly 40,000,000 pounds/year of product. The *704* program with the plant size fixed by the analog optimization could, under some conditions (see Fig. 21, for example), produce more product, hence greater percentage return, than the specified 40,000,000 pounds/year.

The analog and *704* computers yielded flow rates which were within 2% of each other under identical conditions.

```
C     SUBROUTINE FOR REACTOR
C     SET EPSILON TO A MINUS NUMBER TO RETURN STEADY STATE VALUES,
      SUBROUTINE REACTR(VR,TR,CK,EPS)
      DIMENSION FA(6),FB(6),FL(6),FR(6),CK(2,3),RK(6,4),TEM(6),FRDEL(6)
      COMMON FA,FB,FL,FR
   50 CK1 =CK(1,1)*EXPF(-CK(2,1)/(TR+460.))
   60 CK2 =CK(1,2)*EXPF(-CK(2,2)/(TR+460.))
   70 CK3 =CK(1,3)*EXPF(-CK(2,3)/(TR+460.))
   75 ICT= 1
   80 IF(EPS) 120,90,100
   90 PAUSE 11111
  100 ISW = 1
  110 GO TO 140
  120 ISW = 2
  130 EPS = -EPS
  140 DO 150 I = 1,6
  150 TEM(I)=FR(I)
  160 DO 380 I = 1,4
  161 TEMP = TEM(1)+TEM(2)+TEM(3)+TEM(4)+TEM(5)+TEM(6)
  165 TEMQ = TEMP**2
  162 IF(TEMP) 163,163,170
  163 TEMP = 1.0
  170 CON1 =CK1*TEM(1)*TEM(2)/TEMQ
  180 CON2 =CK2*TEM(2)*TEM(3)/TEMQ
  190 CON3 =CK3*TEM(3)*TEM(6)/TEMQ
  195 EEPS = EPS*TEMP
  200 RK(1,I) =EEPS*((FA(1)+FB(1)+FL(1)-TEM(1))/VR-CON1)
  210 RK(2,I) =EEPS*((FA(2)+FB(2)+FL(2)-TEM(2))/VR-CON1-CON2)
  220 RK(3,I) =EEPS*((FA(3)+FB(3)+FL(3)-TEM(3))/VR+2.*CON1-2.*CON2-CON3)
  230 RK(4,I) =EEPS*((FA(4)+FB(4)+FL(4)-TEM(4))/VR+2.*CON2)
  240 RK(5,I) =EEPS*((FA(5)+FB(5)+FL(5)-TEM(5))/VR+1.5*CON3)
      RK(6,I) =EEPS*((FA(6)+FB(6)+FL(6)-TEM(6))/VR+CON2-0.5CON3)
  260 GO TO (270,270,330,360),I
  270 DO 280 J=1,6
  280 TEM(J) = FR(J) + 0.5*RK(J,I)
  290 GO TO 380
  330 DO 340 J =1,6
  340 TEM(J) = FR(J) + RK(J,I)
  350 GO TO 380
  360 DO 375 J=1,6
  370 FRDEL(J) = (RK(J,1)+2.0*RK(J,2)+2.0*RK(J,3)+RK(J,4))/6.0
  375 FR(J) = FR(J) + FRDEL(J)
  380 CONTINUE
  390 GO TO (400,410), ISW
  400 RETURN
  410 SUM = 0.0
  411 IF(ICT-2000) 419,419,412
  412 PAUSE 33333
  413 IF(SENSE SWITCH 1) 400,414)
  414 ICT = 0
  419 ICT = ICT+1
  420 DO 430 I=1,6
  430 SUM = SUM + FRDEL(I)**2
  440 IF(SUM- 0.00000001*TEMQ) 450,450,460
  450 ISW = 1
  460 GO TO 140
      END(2,2,2,2,2)
```

Fig. 30. Reactor subroutine

```
      SUBROUTINE SSSEP (CMPL,ZED,VR,SARE,RINVS)
      DIMENSION FA(6),FB(6),FL(6),FR(6),FS(6),FD(6)
      COMMON FA,FB,FL,FR,FS,FD,SG,SP,SD,PRODRT,PCTRET
      DO 700 I=1,4
  700 FS(I) = FR(I)
      FS(5) = 0.
      FS(6) = CMPL*FS(4)
      SL= 0.
      SR= 0.
      SD= 0.
      DO 800 I=1,6
      FL(I) = FS(I)/(ZED + 1.)
      FD(I) = FL(I)*ZED
      SL=  SL + FL(I)
      SR=  SR + FR(I)
  800 SD = SD + FD(I)
      SP = FR(6) - FS(6)
      SG = FR(5)
      IF(SENSE SWITCH 6)600,500
  500 ZP = MIN1F(SP,4763.)
      GO TO 400
  600 ZP = SP
  400 PRODRT =  0.30*ZP-FA(1)*.02-FB(2)*.03-SG*.01+SD*.0068
      PCTRET = (100.*(8400.*PRODRT-2.22*SR-1041.6*(.30*ZP+.0068*SD))/2780
     1000. -(276,000*100/2,780,000)
      RETURN
      END(2,2,2,2,2,)
```

Fig. 31. Separator and economic equation subroutine

```
VR = 4,640  T = 180    1/Z+1 = 0.55

FLOW RATE OF TAR      =    3.71418E 03
FLOW RATE OF PRODUCT =    4.76481E 03
FLOW RATE OF WASTE    =    3.93695E 04
PRODUCT RETURN        =    3.68971E 02
PERCENT RETURN        =    3.02560E 01


   FA              FB              FL           FR           FS           FD
0.14500E 05   0.              0.63913E 04  0.11621E 05  0.11621E 05  0.52292E 04
0.            0.33350E 05     0.20259E 05  0.36834E 05  0.36834E 05  0.16575E 05
0.            0.              0.12928E 04  0.23506E 04  0.23506E 04  0.10578E 04
0.            0.              0.18341E 05  0.33348E 05  0.33348E 05  0.15006E 05
0.            0.              0.           0.37142E 04  0.           0.
0.            0.              0.18341E 04  0.80996E 04  0.33348E 04  0.15006E 04
```

Fig. 32. Sample format of output, optimum case

Analog Wiring Diagram

The analog wiring diagram is shown in Fig. 28. Although the steady-state equations are a set of algebraic nonlinear equations, one will note that integrators are used in the reactor and reboiler sections. These are used for the elimination of noise within the many loops of the circuit. Physically, the equations as represented in the wiring diagram describe the start-up of a plant simpler than the one reported, but having the same steady-state values. Reactant *A* is initially in all lines. Reactant *B* is added at the time of operation. Note that the reported economic equations 5 and 6 were calculated on the analog as:

$$\frac{0.6V_R}{F_A}(\%\ \text{return}+10)=220.8\frac{F_P}{F_A}-16.8-$$

$$25.2\frac{F_B}{F_A}-0.84\frac{F_G}{F_A}-0.222\frac{F_R}{F_A}+5.0\frac{F_D}{F_A}$$

which is equivalent to equations 5 and 6 when $V_R=4{,}640$ pounds.

All flow rates and the reactor volume are presented as ratios of the reactant stream, F_A. This procedure was followed in order to eliminate a major rescaling of the problem as process parameters were changed over wide ranges.

Table II gives the appropriate potentiometer settings. Approximately 30 seconds are required to reach steady state. The time constant of the integrators was 1 second.

Fortran II Program

The IBM FORTRAN II program is illustrated in Figs. 29 through 32. The steady-state nonlinear equations are solved through numerical integration of differential equations similar to the technique used on the analog computer.

Fig. 29 shows one of the main programs used. The particular one shown gives the steady-state solution for one set of parameters. Other programs (not shown) printed the steady-state solutions to all combinations of incremental changes in the parameters. This latter program was used to calculate Figs. 15 through 18, for instance. No sophisticated optimization techniques (such as the Box technique) were used in the digital program, because the region of the optimum had been well established on the analog computer.

Fig. 30 gives the subroutine for the reactor calculations. The subroutine calculates an incremental change in reactor concentrations and flow rates or, alternately, returns steady-state values for a given input. This second feature, with some modification, will probably be used in the dynamic analysis of the reactor and was not used in the optimization. The numerical integration procedure used is a fourth-degree Runge-Kutta technique.

The steady-state separation of flow streams and the economic equations are solved in one subroutine shown in Fig. 31.

Fig. 32 illustrates the output format for the optimum case.

Appendix X defines the FORTRAN variables in terms of variables listed in the Nomenclature and also the functions of the sense switches.

The approximate running time for the program shown was: 15 seconds, read in; 12 seconds, calculation; 15 seconds, print out.

Appendix X. Fortran Variables

CK(1, I)=A_i
CK(2, I)=B_i
CK1=k_1
CK2=k_2
CK3=k_3
CMPL=P_S/E_S (=0.10)
CON1=$k_1A_RB_R$
CON2=$k_2B_RC_R$
CON3=$k_3C_RP_R$
EEPS=$F_R\times$EPS
EPS=incremental step in time (=0.01 V_R/F_A is satisfactory)
FA(1)=F_A (FA(2) etc., are zero)
FB(2)=F_B (FB(1), FB(3) etc., are zero)
FD(I)=weight rate of flow of component *i* in discard stream. FD(1)=F_DA_S, for example
FL(I)=weight rate of flow of component *i* in recycle
FR(I)=weight rate of flow of component *i* from reactor
FRDEL=incremental change in reactor flow rate
FS(I)=weight rate of flow of component *i* from distillation column
I=subscript; when applied to fluid streams,
  1 refers to component *A*
  2 refers to component *B*
  3 refers to component *C*
  4 refers to component *E*
  5 refers to component *G*
  6 refers to component *P*
ICT=iteration counter
PCTPRE=percentage return of previous iteration
PCTRET=per-cent return
PRODRT=M
RINVS=plant investment cost (not used in program shown)
RK=Runge-Kutta variable
SARE=SARE costs (not used in program shown)
SD=F_D
SG=F_G
SL=F_L
SP=F_P
SR=F_R
SUM=sum of squares of incremental flow rate changes
TEM, TEMP, TEMQ=temporary variable
TR=T_R in F
VR=V_R
ZED=F_D/F_L
ZP=portion of F_P available for sale

Sense Switches (On)	Function
1	.....Resets ICT to zero
2	.....Prints FR(I) at each iteration
3	.....Prints complete output at each iteration
6	.....Calculates economics based on unlimited sale of *P*

Time Response Constant and Transfer Function of Magnetic Amplifiers

E. W. MANTEUFFEL
ASSOCIATE MEMBER AIEE

This paper is an additional contribution to the general theory of magnetic amplifiers and deals with the relationship of the time response constant to the actual transfer function of such amplifiers. Under certain conditions this constant is not the same as that generally understood to be the time constant of an exponential function. In presenting the information given in this paper the author has obtained permission from Dr. Buchhold to review certain basic relationships, contained in Dr. Buchhold's report.

SEVERAL years ago Dr. T. A. Buchhold of the General Electric Company, while still associated with Redstone Arsenal and later with Ford Instrument Company, made theoretical investigations about the transient performance of magnetic amplifiers. In the course of his work the concept of the specific dynamic resistance of square loop core materials and, directly related to it, the concept of the dynamic resistance of magnetic amplifiers was established as a convenient means to express mathematically the performance of such amplifiers. Although Dr. Buchhold presented his findings in an internal company report, his work has unfortunately never been published as a paper. This situation causes certain difficulties in presenting a paper which uses the foregoing concept without at the same time referring in at least some detail to Dr. Buchhold's original work.

Paper 60-875, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting, Atlantic City, N. J., June 19–24, 1960. Manuscript submitted June 16, 1959; made available for printing May 23, 1960.

E. W. MANTEUFFEL is with the General Electric Company Advanced Electronics Center at Cornell University, Ithaca, N. Y.

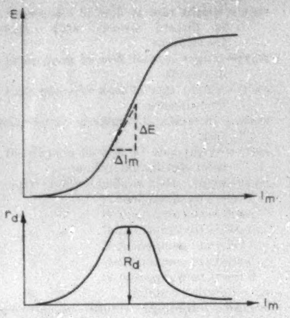


Fig. 1. Dynamic core characteristic

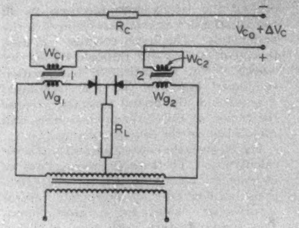


Fig. 2. Center-tap amplifier

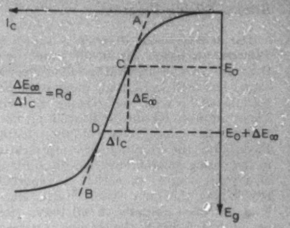


Fig. 3. Dynamic amplifier characteristic E_g versus I_c

Dynamic Resistance, Time Response Constant, and Voltage Gain

Assume that a magnetic core of square loop core material with cross section A and mean magnetic path length l_m is wound with a winding of N turns and energized from an a-c source. If one determines by known methods the relationship between the half-cyclic average value of electromotive force E and the average value of magnetizing current I_m, the dynamic characteristic illustrated in Fig. 1 is obtained. For the following consideration it is of minor importance how this characteristic is determined. It may, for example, be measured with the well-known constant-current flux reset method. Buchhold has shown that in the case of square loop core materials, it also can be measured by simply increasing the a-c supply voltage from zero to some maximum value and by recording the half-cyclic average values of magnetizing current and electromotive force. The author and several of his associates have obtained good results with the latter method but the scope of this report does not allow going into all the details.

The slope $\Delta E/\Delta I_m$ at any point of the dynamic characteristic expresses a certain dynamic resistance r_d which is plotted in the lower portion of Fig. 1. Since in most practical applications magnetic amplifiers are operated within the linear region of their characteristic, the maximum slope at the center portion of the dynamic characteristic is simply called the dynamic resistance R_d. Since this value R_d is related to particular core dimensions and number of turns, it is preferable to express it in per-unit terms as

$$R_{d_0} = \frac{\Delta E_0}{\Delta I_{m_0}}[\text{ohms/cm (per centimeter)}] \qquad (1)$$

where

ΔE_0 = change of half-cyclic average value of emf (electromotive force) per turn per cm² cross section

and

ΔI_{m_0} = change of half-cyclic average value of magnetizing ampere-turns per cm magnetic path length

This value R_{d_0}, called the specific dynamic resistance, is a significant parameter of a given core. Its value depends on the kind of core material, the supply frequency, the ratio of inside to outside diameter of the core, and also to a certain degree on temperature. The actual dynamic resistance R_d is then related to R_{d_0} as follows:

$$R_d = R_{d_0} \cdot N^2 \cdot \frac{A}{l_m} \qquad (2)$$

Because of secondary effects, the specific dynamic resistance determined from the actual transfer characteristic of a magnetic amplifier is generally not the same as that obtained from core measurements. Deviations result from rectifier leakage, effects of gate limiting, and the actual ratio of control circuit resistance to dynamic resistance. Rectifier leakage is almost negligible for silicon rectifiers. The other effects can be taken into account by correlation factors. For the following discussion it is assumed that the actual dynamic characteristic including those secondary effects is known. This characteristic can always be measured by recording the actual half-cyclic average values of gate emf as a function of control current and referring both values either to the number of control turns N_c or turns of the gate winding N_g.

In order to explain in some detail the influences of dynamic resistance R_d and control circuit resistance R_c on the time response constant T and the specific volt-

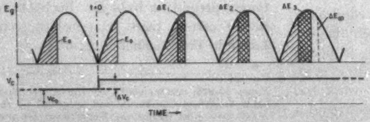
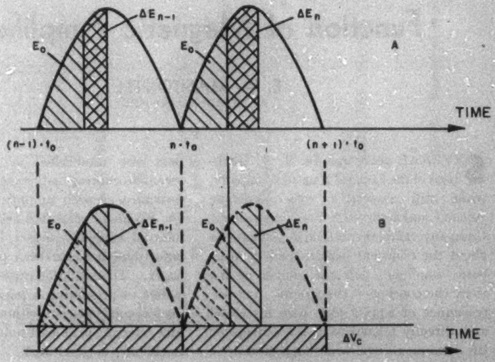


Fig. 4 (above). Gate emf's versus control voltage step function

Fig. 5 (right). Gate emf's and control voltage in two successive half-periods



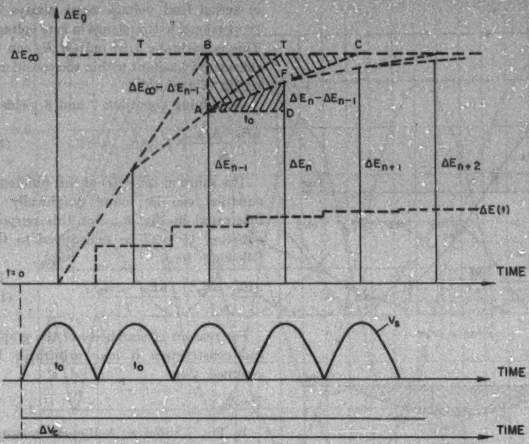


Fig. 6. Graphical construction of function $\frac{\Delta E_n-\Delta E_{n-1}}{t_o}=\frac{\Delta E_\infty-\Delta E_{n-1}}{T}$

age gain G_v, of a magnetic amplifier, the well-known center-tap amplifier is to be investigated. The basic circuit configuration of such an amplifier is illustrated in Fig. 2. An actual dynamic characteristic

$$E_g=f(I_c)$$

as described before is shown in Fig. 3. This characteristic is intentionally plotted within the third quadrant to show better its resemblance to the transfer characteristic of an amplifier. In its center portion the dynamic characteristic is approximated by a straight line AB. For the following considerations the number of control and gate turns may be made equal.

As a starting point it is assumed that a bias voltage V_{c_o} had been previously impressed on the control circuit and that the amplifier is in steady-state operation with a firing angle somewhat smaller than 90 degrees. This condition is indicated as point C in Fig. 3. It is also shown in Fig. 4, where the single-hatched areas indicate the steady-state gate emf, E_0. Let us now assume that at time $t=0$, at the start of a half cycle, the voltage at the control circuit is increased in step-fashion by an amount ΔV_c as shown in the lower portion of Fig. 4. The result will be, as is generally known, that in following half cycles the gate voltage will become larger by increasing amounts ΔE_1, ΔE_2, $\Delta E_3 \ldots$ (indicated by cross-hatched areas) until finally in the new steady-state operation an increase ΔE_∞ is reached. This value is shown as a dashed line in Fig. 4 and corresponds to point D in Fig. 3. The sequence of events as they occur from half cycle to half cycle will now be described mathematically by considering conditions at the $(n-1)$th and nth half-period after the step voltage ΔV_c was applied. The time will be measured in multiples of time t_o of one half cycle.

$$t_o=\frac{1}{2f}$$

where f is the frequency of the supply source.

The half cycle which starts at the time $(n-1)\ t_o$ will be investigated. During this time interval, the gate electromotive force E_g, which in the course of preceding half cycles has already increased from E_o to $E_o+\Delta E_{n-1}$, may arbitrarily exist at winding W_{g2} of core 2 during its gating half cycle. This condition is shown in Fig. 5(A). This same voltage $E_o+\Delta E_{n-1}$ is also induced in winding W_{c_2} of the control circuit since equal numbers of turns were assumed. Its polarity is such that it adds to the total control voltage $V_{c_o}+\Delta V_c$ and assists in resetting core 1 during this half cycle.

Since as a starting condition, steady-state operation of the amplifier at point C, Fig. 3, was assumed, it is sufficient to consider only variations of the gate voltage under the influence of ΔV_c. Therefore, the change in resulting electromotive force is actually caused by ΔV_c and ΔE_{n-1} as indicated by the solidly hatched area of Fig. 5(B). These areas are considered as half-cyclic average values. The change in average gate voltage of core 1 is designated as ΔE_n since ΔE_n is the change of E_g during the gating half cycle starting at time nt_o. Therefore the following relationship between ΔE_n, ΔE_{n-1}, and ΔV_c must exist:

$$\Delta E_n=\left(\Delta V_{c_{n-1}}+\Delta E_{n-1}\right)\frac{R_d}{R_d+R_c} \tag{3}$$

For a more general consideration it may be assumed that additional external electric feedback is employed. This feedback can be expressed by a factor (1-h_e) associated with ΔE_{n-1}. Equation 3 changes then to

$$\Delta E_n=\left[\Delta V_{c_{n-1}}+(1-h_e)\,\Delta E_{n-1}\right]\frac{R_d}{R_d+R_c} \tag{4}$$

By rearranging terms and observing

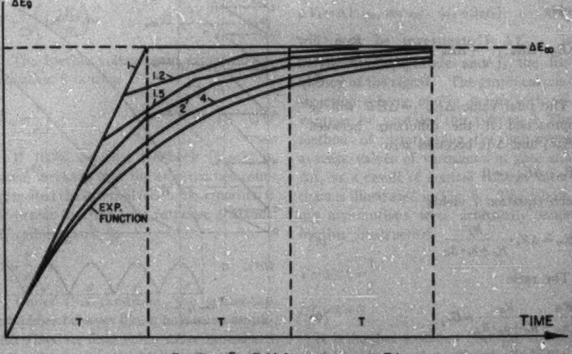


Fig. 7. $\int \Delta E_g(t)$ for various ratios T/t_o

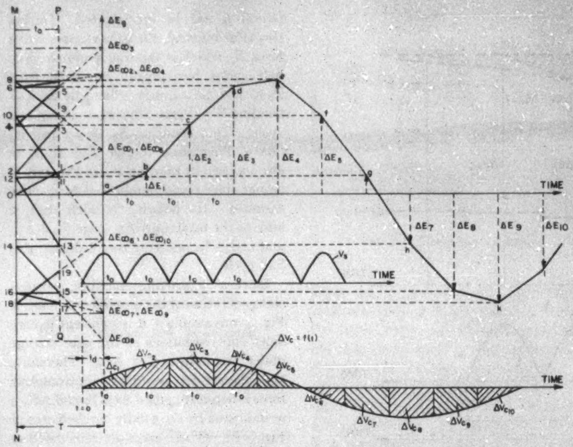

M
P
N
Q
t_0
ΔE_g
$\Delta E_{\infty 3}$
$\Delta E_{\infty 2}, \Delta E_{\infty 4}$
$\Delta E_{\infty 1}, \Delta E_{\infty 5}$
$\Delta E_{\infty 6}, \Delta E_{\infty 10}$
$\Delta E_{\infty 7}, \Delta E_{\infty 9}$
$\Delta E_{\infty 8}$
ΔE_1
ΔE_2
ΔE_3
ΔE_4
ΔE_5
ΔE_7
ΔE_8
ΔE_9
ΔE_{10}
a
b
c
d
e
f
g
h
k
TIME
V_s
$\Delta V_c = f(t)$
ΔV_{c1}
ΔV_{c2}
ΔV_{c3}
ΔV_{c4}
ΔV_{c5}
ΔV_{c6}
ΔV_{c7}
ΔV_{c8}
ΔV_{c9}
ΔV_{c10}
t_d
t = 0
T
0
2
4
6
8
10
12
14
16
18

Fig. 8. Graphical construction of function $\int \Delta E_g(t)$ for sinusoidally varying control voltage

that $\Delta V_{c_{n-1}} = \Delta V_c$ for all half cycles, this equation becomes

$$(\Delta E_n - \Delta E_{n-1}) \frac{R_c + R_d}{R_c + h_e \cdot R_d} + \Delta E_{n-1} = \Delta V_c \frac{R_d}{R_c + h_e \cdot R_d}$$

If the numerator and denominator of the first term are multiplied by $t_0 = 1/2f$ one obtains

$$\frac{\Delta E_n - \Delta E_{n-1}}{t_0} \cdot \frac{R_c + R_d}{R_c + h_e \cdot R_d} \cdot \frac{1}{2f} + \Delta E_{n-1} = \Delta V_c \frac{R_d}{R_c + h_e \cdot R_d} \qquad (5)$$

Making the substitution

$$T = \frac{1}{2f} \cdot \frac{R_c + R_d}{R_c + h_e \cdot R_d} \qquad (6)$$

yields

$$(\Delta E_n - \Delta E_{n-1}) \frac{T}{t_0} + \Delta E_{n-1} = \Delta V_c \frac{R_d}{R_c + h_e \cdot R_d} \qquad (7)$$

The final value $\Delta E_{n-1} = \Delta E \infty$ will be approached if the difference between ΔE_{n-1} and ΔE_n becomes zero

$$\Delta E_n - \Delta E_{n-1} = 0$$

Then equation 7 yields

$$\Delta E_\infty = \Delta V_c \cdot \frac{R_d}{R_c + h_e \cdot R_d} \qquad (8)$$

The ratio

$$\frac{AE_\infty}{V\Delta_c} = \frac{R_d}{R_c + h_e \cdot R_d} = G_{v_s} \qquad (9)$$

is called the specific voltage gain of the amplifier, which is a theoretical value. The actual voltage gain G_v can then be found by taking into account the influence of gate circuit resistance R_g and the turns ratio N_g/N_c which will in general be different from "one."

$$G_v = \frac{\Delta V_L}{\Delta V_c} = \frac{R_d}{R_c + h_e \cdot R_d} \cdot \frac{R_L}{R_L + R_g} \cdot \frac{N_g}{N_c} \qquad (10)$$

The value ΔV_L can without restriction be used in this equation since the change in actual load voltage must always be proportional to the change in gate voltage. This is naturally true only if R_d can be considered constant within the considered region of variations.

Combining equations 7 and 8 yields

$$\frac{\Delta E_n - \Delta E_{n-1}}{t_0} \cdot T + \Delta E_{n-1} = \Delta E_\infty \qquad (11)$$

The solution $\Delta E = f(t)$ of this difference equation can be found graphically as illustrated in Fig. 6. For this purpose equation 11 shall be rearranged to the following form:

$$\frac{\Delta E_n - \Delta E_{n-1}}{t_0} = \frac{\Delta E_\infty - \Delta E_{n-1}}{T} \qquad (12)$$

For reasons of simplicity of the graphical construction it may arbitrarily be assumed that

$$T = 2t_0$$

In Fig. 6 different half-cyclic average values ΔE are plotted at intervals of t_0. They are shown as arrows ΔE_{n-1}, ΔE_n, ΔE_{n+1} . . . and placed at the centers of the corresponding half cycles of the supply voltage which is indicated below as a rectified sine wave. The instant where ΔV_c is applied is shown in correspondence with Fig. 4 at $t = 0$. The final value $\Delta E \infty$ determined by equation 8 is indicated as a dashed horizontal line. The values of ΔE at each half cycle can be found by a simple graphical construction. For example, the value of ΔE_n can be determined by connecting point A of arrow ΔE_{n-1} with point C on the horizontal $\Delta E \infty$ line by a

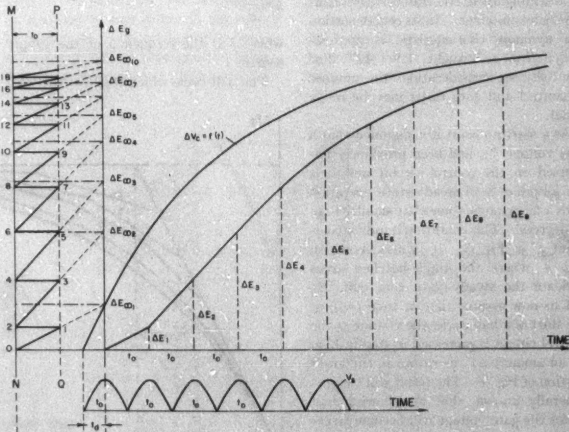

M
P
N
Q
t_0
ΔE_g
$\Delta E_{\infty 1}$
$\Delta E_{\infty 2}$
$\Delta E_{\infty 3}$
$\Delta E_{\infty 4}$
$\Delta E_{\infty 5}$
$\Delta E_{\infty 7}$
$\Delta E_{\infty 10}$
$\Delta V_c = f(t)$
ΔE_1
ΔE_2
ΔE_3
ΔE_4
ΔE_5
ΔE_6
ΔE_7
ΔE_8
ΔE_9
TIME
V_s
t_d
T
0
2
4
6
8
10
12
14
16
18

Fig. 9. Graphical construction of $\int \Delta E_g(t)$ for exponentially rising control voltage

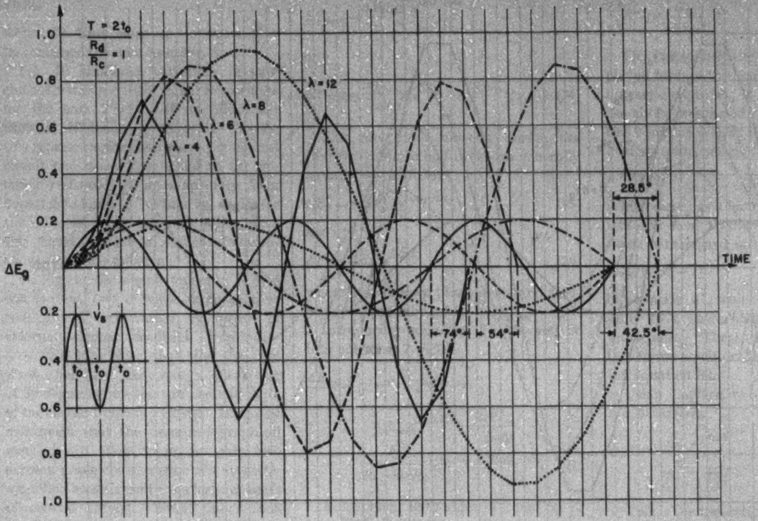


Fig. 10. Variations in gate emf for various control signal frequencies

straight line where point C is time-wise by T apart from ΔE_{n-1}. ΔE_n is then found at point F at the time distance t_0 from ΔE_{n-1}. The correctness of this construction is verified if one considers the two shaded similar triangles ABC and ADF whose ratios of corresponding sides

$$\frac{AB}{BC}=\frac{DF}{AD}$$

are nothing other than an expression of equation 12.

The value

$$T=\frac{1}{2f}\cdot\frac{R_c+R_d}{R_c+h_e\cdot R_e} \tag{6}$$

now becomes meaningful. It is the time response constant of a magnetic amplifier which is not identical with the time constant τ of an exponential function. Agreement between T and τ will be only approached when

$$T>>t_0$$

In Fig. 6, average half-cyclic values ΔE are also indicated as a dashed staircase line. One can show mathematically that the difference equation describing this function is given by

$$\Delta E_n(t)=\Delta E_\infty\left[1-\left(1-\frac{t_0}{T}\right)^n\right] \tag{13}$$

This equation permits the value of ΔE_n to be calculated at any given half cycle n after application of the step signal ΔV_c.

In Fig. 7 several time functions for different values of T/t_0 together with the exponential function are illustrated. The time response constant T is the same for all the curves but the carrier-frequency f and, therefore, t_0 is different. Values of T/t_0 are indicated in this figure by numbers. Points at different half cycles are connected by straight lines for better display, although in reality the function $\Delta E=f(t)$ is a jump-function.

The ratio T/t_0 can in general be changed according to equation 6 by different R_e values or various amounts of external feedback. If no external feedback is employed ($h_e=0$), equation 6 becomes

$$T=\frac{1}{2f}\left(1+\frac{R_d}{R_e}\right) \tag{14}$$

The specific voltage gain expressed by equation 9 is then

$$G_{v_s}=\frac{R_d}{R_c} \tag{15}$$

If 100% external feedback ($h_e=1$) is used, meaning that the external has compensated the internal feedback, equation 6 indicates that the time response constant T reduces now to

$$T=\frac{1}{2f}=t_0 \tag{16}$$

Under this condition, the center-tap amplifier behaves like a half-wave amplifier since its final value ΔE_∞ is reached after the time $t=t_0$, which is also apparent from equation 11. The voltage gain, equation 9, yields then

$$G_{v_s}=\frac{R_d}{R_d+R_c}=G_{v_{s0}} \tag{17}$$

If the performance of a simple half-wave amplifier is investigated, the same equation 17 will be obtained for its specific voltage gain.

Transfer Function of Magnetic Amplifier

The response of the discussed center-tap amplifier to a sinusoidally varying control signal voltage

$$\Delta V_c=\Delta V_{c_{\max}}\sin\omega_s t \quad (\omega_s=2\pi f_s) \tag{18}$$

will now be investigated. $\Delta V_{c_{\max}}$ is hereby the amplitude and f_s the frequency of the signal. The graphical construction scheme of Fig. 6 can with advantage be used for this purpose. The method of determining the half-cyclic average values of variations in gate emf ΔE_g as a result of control voltage variations is illustrated in Fig. 8. The following assumptions were arbitrarily made for this construction.

$$T=2\times t_0=\frac{1}{f}$$

$$f/f_s=\lambda=5$$

It is further assumed that the control signal starts at time $t=0$ at the beginning

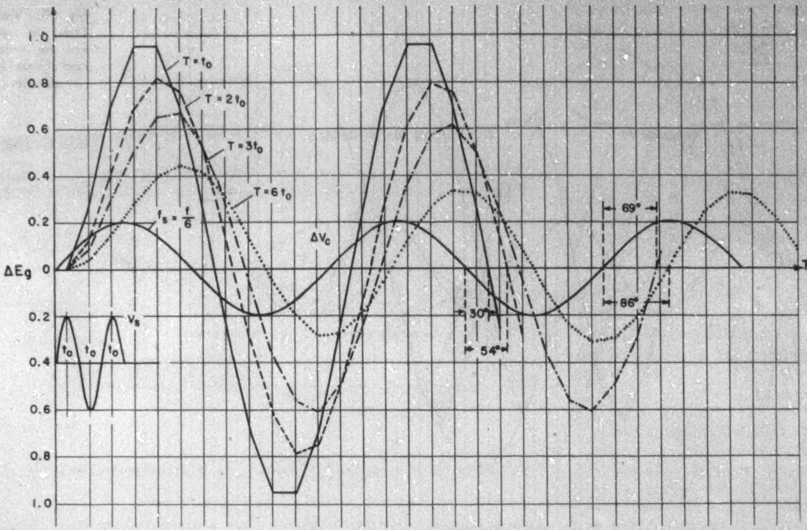


Fig. 11. Variations in gate emf at constant signal frequency for various ratios T/t₀

of a half-period t_o. The control voltage is shown as curve

$$\Delta V_c = f(t)$$

During each half-cycle t_o different half-cyclic average values of control voltages ΔV_{c1}, ΔV_{c2}, ΔV_{c3}, ... act on the control circuit of the magnetic amplifier. These voltages are shown as differently hatched areas in Fig. 8. If each of these voltages existed for an indefinitely long time, they would result in final gate emf's $\Delta E_{\infty 1}$, $\Delta E_{\infty 2}$, $\Delta E_{\infty 3}$, ... as determined by equation 8.

The particular values ΔV_{c_n} for each half cycle are given by

$$\Delta V_{c_n} = \frac{\lambda}{\pi}\left[\cos\frac{(n-1)\pi}{\lambda} - \cos\frac{n\pi}{\lambda}\right]\Delta V_{c_{max}} \tag{19}$$

The values of $\Delta E_{\infty n}$ are then

$$\Delta E_{\infty n} = \Delta V_{c_n} \cdot G_{v_s} \tag{20}$$

They are indicted as dot-dashed horizontal lines on the left of Fig. 8. For the construction a line *M-N* is drawn in parallel to the ordinate axis at a distance *T*. A second parallel line *P-Q* at a distance t_o from line *M-N* is also indicated. The trace of lines 0-1-2-3-4- ... 18-19 shows how values ΔE_1, ΔE_2, ΔE_3 ... can be successively determined. Starting at point 0 a line is drawn connecting 0 with $\Delta E_{\infty 1}$ at the ordinate axis. The point 1, where this line intersects with line *P-Q*, determines the value ΔE_1. Returning now in a horizontal direction to point 2, a line connecting 2 with $\Delta E_{\infty 2}$ is drawn and ΔE_2 is then found at point 3. The construction continues then in the same fashion. The values ΔE_1, ΔE_2, ΔE_3, ... are plotted at the center of corresponding half cycles. The solid line *a-b-c-d* ... *k*-1 ... then represents the variation of the gate emf ΔE_g on an average basis. On this basis, a median dead time lag

$$t_d = \frac{t_o}{2} = \frac{1}{4f} \tag{21}$$

exists. The construction scheme of Fig. 8 can be used for any kind of function of control voltage versus time. Fig. 9 shows as an example the response when an exponentially rising control voltage

$$\Delta V_c = \Delta V_{c_\infty}[1 - \epsilon^{-t/\tau}]$$

is applied to the amplifier. For the construction, τ was arbitrarily chosen to

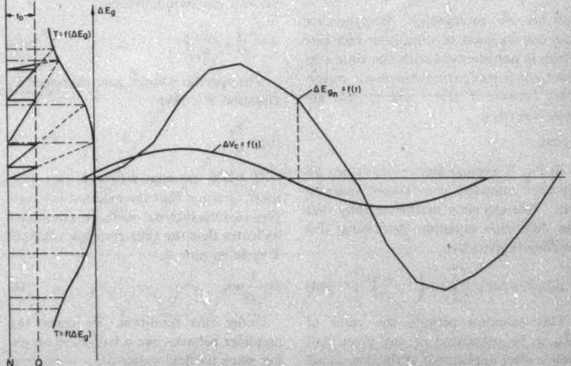


Fig. 12. Nonlinear response to sinusoidally varying control voltage

$\tau = 3t_o$

while the time response constant $T = 2t_o$ remained the same as in Fig. 8.

Fig. 10 illustrates variations in gate emf for various control signal frequencies for the case $T = 2t_o$. If no additional external feedback is assumed ($h_e = 0$), this represents the case where the control circuit resistance R_c is equal to the dynamic resistance R_d as is apparent from equations 14 and 15. Phase lag angles for the different frequencies are indicated in this figure. Relative values of amplitudes of ΔE_g are apparent.

In Fig. 11 the signal frequency is held constant at $f = f/6$ while response time constants T are varied by means of external electric feedback from $T = t_o$ to $T = 6t_o$. Note that, for the case $T = t_o$, half-cycle response, a phase lag of 30 degrees equivalent to one half cycle of the carrier frequency exists. It may be mentioned that the described graphical method will allow taking roughly into account nonlinear operation of a magnetic amplifier, as will result from large changes of control voltage. Since, under such conditions, the dynamic resistance and the specific voltage gain are subject to variations, the distance T of Fig. 8 therefore has to be gradually changed as a function of ΔE_g.

Also the values $\Delta E_{\omega n}$ must be determined in correspondence with the dynamic characteristic of the amplifier. The general scheme of graphical construction for this case is shown in Fig. 12, where the reference values $\Delta E_{\omega 1}$, $\Delta E_{\omega 2}$ ... are now located on the curve $T = f(\Delta E_g)$.

The response of the magnetic amplifier to a sinusoidally varying control signal voltage will now be investigated mathematically as follows: It was shown before that the relationship between gate emf's at succeeding half cycles is expressed by

$$(\Delta E_n - \Delta E_{n-1})\frac{T}{t_o} + \Delta E_{n-1} = \Delta V_{c_{n-1}} \cdot \frac{R_d}{R_c + h_e \cdot R_d}$$

from equation 7.

The control signal voltage is

$$\Delta V_c = \Delta V_{c_{max}} \sin \omega_s t$$

as expressed in equation 18.

Assuming an arbitrary phase angle of ΔV_c, one can also write

$$\Delta V_c = \Delta V_{c_{max}} \epsilon^{j\omega_s t} \tag{22}$$

The average value of ΔV_c during the $(n-1)$th half-cycle is then

$$\Delta V_{c_{n-1}} = \frac{1}{t_o}\int_{(n-1)t_o}^{nt_o} \Delta V_{c_{max}} \epsilon^{j\omega_s t}\, dt \tag{23}$$

$$\Delta V_{c_{n-1}} = \frac{\Delta V_{c_{max}}}{t_o}\left[\frac{\epsilon^{j\omega_s t}}{j\omega_s}\right]_{(n-1)t_o}^{nt_o}$$

or

$$\Delta V_{c_{n-1}} = \frac{\Delta V_{c_{max}}}{j\omega_s t_o}\left[\epsilon^{j\omega_s t_o} - \epsilon^{j\omega_s (n-1)t_o}\right] \tag{24}$$

The gate emf's existing during the $(n-1)$th and nth half-cycle can generally be expressed as follows:

$$\Delta E_{n-1} = \Delta E_{max} \epsilon^{j\omega_s (n-1)t_o} \tag{25}$$

and

$$\Delta E_n = \Delta E_{max} \epsilon^{j\omega_s n t_o} \tag{26}$$

where an unknown phase angle and magnitude are attributed to ΔE_{max}.

Introducing equations 24, 25, and 26 into equation 7 yields:

$$\frac{\Delta E_{max} T}{t_o}\left[\epsilon^{j\omega_s n t_o} - \epsilon^{j\omega_s (n-1)t_o}\right] + \Delta E_{max} \epsilon^{j\omega_s (n-1)t_o} = \frac{\Delta V_{c_{max}}}{j\omega_s t_o} \frac{R_d}{R_c + h_e \cdot R_d}\left[\epsilon^{j\omega_s n t_o} - \epsilon^{j\omega_s (n-1)t_o}\right] \tag{27}$$

Multiplying both sides of equation 27 by $\epsilon^{-j\omega_s (n-1)t_o}$ yields

$$\frac{\Delta E_{max} T}{t_o} \epsilon^{j\omega_s t_o} - 1 + \Delta E_{max} = \frac{\Delta V_{c_{max}}}{j\omega_s t_o} \frac{R_d}{R_c + h_e \cdot R_d}\left[\epsilon^{j\omega_s t_o} - 1\right] \tag{28}$$

By rearranging terms:

$$\frac{\Delta E_{max} T}{t_o}\left[\epsilon^{j\omega_s t_o} - 1 + \frac{t_o}{T}\right] = \frac{\Delta V_{c_{max}}}{j\omega_s t_o} \frac{R_d}{R_c + h_e \cdot R_d}\left(\epsilon^{j\omega_s t_o} - 1\right)$$

or

$$\frac{\Delta E_{max}}{\Delta V_{c_{max}}} = \frac{R_d}{R_c + h_e \cdot R_d} \frac{1}{j\omega_s T} \frac{\epsilon^{j\omega_s t_o} - 1}{\epsilon^{j\omega_s t_o} - 1 + \frac{t_o}{T}} \tag{29}$$

$$\frac{\Delta E_{max}}{\Delta V_{c_{max}}} = G_{v_s} \frac{1}{\frac{j\omega_s t_o}{\epsilon^{j\omega_s t_o} - 1} + j\omega_s T} \tag{30}$$

Equation 30 has to be multiplied by $\epsilon^{-j\omega_s t_o/2}$ in order to account for the median transportation lag $t_d = t_o/2$; see equation 21. Therefore one obtains

$$\frac{\Delta E_{max}}{\Delta V_{c_{max}}} = G_{v_s} \frac{\epsilon^{-j\omega_s t_o/2}}{\frac{j\omega_s t_o}{\epsilon^{j\omega_s t_o} - 1} + j\omega_s T} \tag{31}$$

By applying l'Hospital's rule to the first term of the denominator, one finds that

$$lim \frac{j\omega_s t_o}{\epsilon^{j\omega_s t_o} - 1} = 1 \text{ for } \omega_s t_o \rightarrow 0$$

This indicates that for the condition of small values of the ratio f_s/f equation 31 can be replaced by

$$\frac{\Delta E_{max}}{\Delta V_{c_{max}}} \cong \frac{G_{v_s}}{1 + j\omega_s T}. \tag{32}$$

This is obviously the conventionally employed form of the trausfer function for a magnetic amplifier; it is valid only with the above-mentioned restrictions. Often this equation is used in a form where the median transportation lag

$$\epsilon^{-j\omega_s t_o/2}$$

is introduced as a factor into the numerator. As soon as this term gains importance with regard to phase lag, the term

$$j\omega_s t_o / (\epsilon^{j\omega_s t_o} - 1)$$

of equation 31, which contributes essentially to a phase lead, must gain equal importance. Therefore, it is more correct to use equation 31 instead of modifying equation 32.

Using Laplace notation, equation 31 can be written

$$G_{th}(s) = \frac{G_{v_s} \epsilon^{-st_o/2}}{\frac{st_o}{\epsilon^{st_o} - 1} + sT} \tag{33}$$

This is the theoretical transfer function of a magnetic amplifier, which relates variations in control voltage to variations in gate emf on an equal turns basis.

The over-all transfer function which takes into account turns ratio N_g/N_c and internal voltage drops is then determined by:

$$G(s) = G_{v_s} \frac{N_g}{N_c} \frac{R_L}{R_L + R_g} \frac{\epsilon^{-st_o/2}}{\frac{st_o}{\epsilon^{st_o} - 1} + sT} \tag{34}$$

Equation 31 can also be written in a somewhat different fashion if one remembers that the ratio G_{v_s}/T amounts according to equations 6 and 9 to

$$\frac{G_{v_s}}{T} = 2f \frac{R_d}{R_c + R_d} \tag{35}$$

where

$$\frac{R_d}{R_c + R_d} = G_{v_{s0}}$$

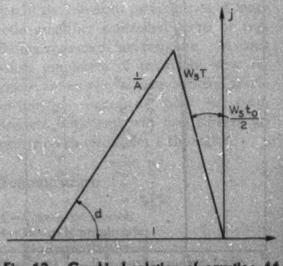


Fig. 13. Graphical solution of equation 44

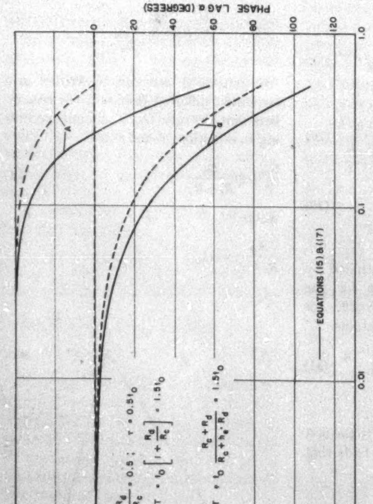


Fig. 14. Attenuation and phase angle versus ratio f_s/f for $R_d/R_c=0.5$

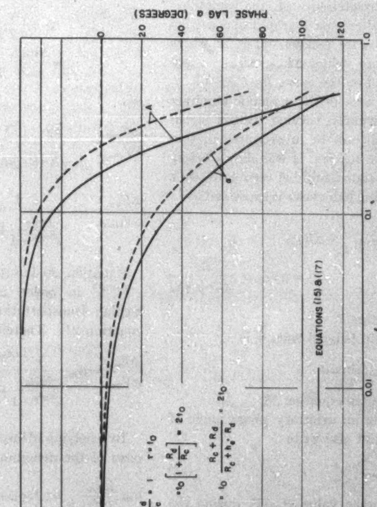


Fig. 15. Attenuation and phase angle versus ratio f_s/f for $R_d/R_c=1$

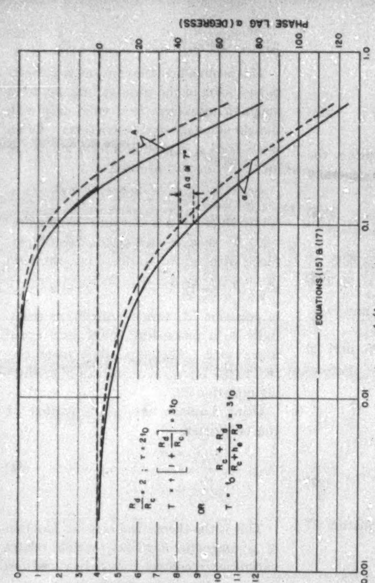


Fig. 16. Attenuation and phase angle versus ratio f_s/f for $R_d/R_c=2$

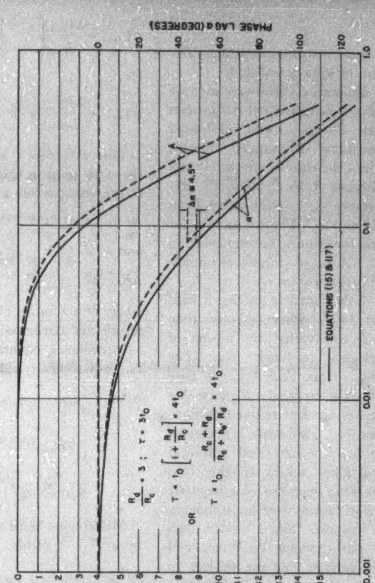


Fig. 17. Attenuation and phase angle versus ratio f_s/f for $R_d/R_c=3$

equation 17 is the specific voltage gain of an equivalent half-wave amplifier. Combining equations 35 and 31 yields:

$$\frac{\Delta E_{max}}{\Delta V_{c_{max}}}=2f\frac{R_d}{R_c+R_d}\frac{\epsilon^{-j\omega_s t_0/2}}{j\omega_s+\dfrac{j\omega_s}{\epsilon^{j\omega_s t_0}-1}\dfrac{t_0}{T}} \tag{36}$$

Letting $T=t_0$ (half-cycle response) equation 36 after rearrangement of terms becomes

$$\frac{\Delta E_{max}}{\Delta V_{c_{max}}}=\frac{R_d}{R_c+R_d}\frac{\epsilon^{j\omega_s t_0/2}-\epsilon^{-j\omega_s t_0/2}}{j\omega_s t_0\cdot\epsilon^{j\omega_s t_0}} \tag{37}$$

or in different form

$$\frac{\Delta E_{max}}{\Delta V_{c_{max}}}=\frac{R_d}{R_c+R_d}\frac{2\sin\omega_s\dfrac{t_0}{2}}{\omega_s t_0}\epsilon^{-j\omega_s t_0} \tag{38}$$

Equation 38 indicates that the phase lag of the half-cycle response amplifier amounts to

$$\alpha=\omega_s t_0=\pi\frac{f_s}{f}\text{ (radians)}$$

which is a generally known fact. The same result was, as mentioned earlier, obtained by graphical means shown in Fig. 11.

In the existing literature the "time constant" of a magnetic amplifier is often expressed as follows:

$$\tau=\frac{1}{2f}\frac{R_0}{R_c}\frac{N_c}{N_g}\frac{\Delta I_L}{\Delta I_c} \tag{39}$$

where

$$R_0=R_L+R_g$$

Letting

$$N_c/N_g=1\text{ and }\Delta I_L(R_L+R_g)=\Delta E_g$$

$$\tau=\frac{1}{2f}\frac{1}{R_c}\frac{\Delta E_g}{\Delta I_c}$$

but

$$\frac{\Delta E_g}{\Delta I_c}=R_d\text{ (as defined in this report)}$$

Therefore

$$\tau=\frac{1}{2f}\frac{R_d}{R_c} \tag{40}$$

Comparing equation 40 with equation 14, one realizes that the term 1 is obviously missing. Therefore equation 40 should furnish wrong results, especially in cases where R_d/R_c is not considerably larger than 1. In many practical applications this ratio should be made at least comparable to 1 since the time response constant must be kept at a suitable low value. In addition, it is known that a large R_d/R_c ratio results in quite a considerable deviation of the transfer-characteristic from linearity. From a theoretical viewpoint, as R_d/R_c becomes smaller equation 40 indicates a time constant approaching zero for very large R_c values. This cannot be true, since the minimum response time obtainable should still be the time of one half cycle, namely $t_0=1/2f$. The more correct relationship is therefore given by equation 14 and, in case of external electric feedback, by equation 6.

In addition, the conventionally used transfer function

$$G(\omega_s)=G_v\frac{\epsilon^{-j\omega_s t_0/2}}{1+j\omega_s\tau}\quad\left[\tau=\frac{1}{2f}\frac{R_d}{R_c}\right] \tag{41}$$

should be, at least in all cases where R_d/R_c is comparable to 1, replaced by the function

$$G(\omega_s)=G_v\frac{\epsilon^{-j\omega_s t_0/2}}{\dfrac{j\omega_s t_0}{\epsilon^{j\omega_s t_0}-1}+j\omega_s T}\quad\left[T=\frac{1}{2f}\left(1+\frac{R_d}{R_c}\right)\right] \tag{42}$$

A study of the first term of the denominator of equation 42 will reveal that

$$\epsilon^{-j\omega_s t_0/2}\cong\frac{j\omega_s t_0}{\epsilon^{j\omega_s t_0}-1} \tag{43}$$

can be substituted for all values of $\omega_s t_0$ smaller than $\pi/4$. Introducing equation 43 in equation 42 and rearranging terms gives

$$G(\omega_s)\cong G_v\frac{1}{1+j\omega_s T\cdot\epsilon^{j\omega_s t_0/2}}\quad(\omega_s t_0<\pi/4) \tag{44}$$

Phase lag angle α and relative amplitude A of this function can be determined by a simple graphical construction as illustrated in Fig. 13, which is self-explanatory. In order to demonstrate actual differences in results obtained from equation 41, and equations 44 or 42, respectively, calculations of phase lag angles α and per-unit amplitudes A were made for a variety of conditions assuming G_v to be unity. Calculations were made for values R_d/R_c ranging from 0.5 to 3.0 and various ratios f_s/f ranging from 0.01 to 0.5. Results are illustrated in Figs. 14 to 17. As was to be expected, differences in results obtained by equation 41 versus equations 42 or 44 respectively, become smaller with increasing values R_d/R_c.

It is of interest to note that in the cases of larger R_d/R_c values an almost constant additional phase lag $\Delta\alpha$ can be observed above a certain signal-to-carrier-frequency ratio, if one compares the more accurate values of equations 42 and 44 with those of equation 41. For the case $R_d/R_c=2$, this angle $\Delta\alpha$ amounts to about 7 degrees for $f_s/f>0.08$, while for $R_d/R_c=3$ it is approximately 4.5 degrees for $f_s/f>0.05$. Such additional phase lags may well be taken into account in the design of certain fast responding control or regulating systems, especially where two or three magnetic amplifiers are used in cascade. Also, in systems where problems of structural resonance occur these phase lags should be considered; see Figs. 16 and 17.

Table I. Values from Fig. 10

$T=2t_0$ f_s/f	From Equation 44		From Graph	
	A	α	A	α
1/12	0.943	29	0.930	28.5
1/8	0.870	42.5	0.860	42.5
1/6	0.804	54.2	0.790	54
1/4	0.665	74	0.660	74

Table II. Values from Fig. 11

$f_s/f=1/6$ T	From Equation 44		From Graph	
	A	α	A	α
t_0	0.990	30	0.955	30
$2t_0$	0.804	54.2	0.790	54
$3t_0$	0.614	69	0.610	69
$6t_0$	0.328	86.5	0.320	86

In order to have a comparison between results obtained by the previous discussed graphical method and results from equation 44, values of A and α were also calculated for the conditions illustrated in Figs. 10 and 11. These calculated values and values taken from those figures are given in Tables I and II.

It may be stated that in the case of large R_d/R_c values, if external electric feedback is employed, equation 6 for the time response constant simply reduces to:

$$T\cong\frac{t_0}{h_e}\cong\frac{1}{2fh_e}$$

This occurs quite often in practice, if fast response is combined with reasonable power gain.

Conclusions

Equations and graphical methods presented will permit more accurate prediction of the transient performance of magnetic amplifier controlled servo or regulating systems of fast response. In cases where amplifiers of relatively slow response are employed, conventional transfer-function equations may be used. Readers interested will find further information in references 1 through 4.

Nomenclature

A = Core cross section
A = Per-unit amplitude
E = Electromotive force
E_g = Gate winding electromotive force

ΔE_e = Change of half-cylic average emf per turn per cm³
ΔE_n = Total change of half-cyclic average value gate emf at n^{th} half-cycle
ΔE_{n-1} = Total change of half-cyclic average value gate emf at $(n-1)^{th}$ half-cycle
ΔE_∞ = Final change in gate emf for defined change of control voltage
f = Carrier frequency
f_s = Signal frequency
G_{vs} = Specific voltage gain
G_v = Total voltage gain
G = Transfer function
h_e = External electric feedback factor
I_c = Control current
I_L = Load current
ΔI_{m_0} = Change of half-cyclic average magnetizing ampere-turns per cm path length
I_m = Magnetizing current
l_m = Mean magnetic path length
N = Number of turns
N_g = Number of turns of gate winding
N_c = Number of turns of control winding
R_d = Dynamic resistance
R_{d_0} = Specific dynamic resistance
R_c = Control circuit resistance
R_L = Load resistance
R_g = Gate circuit resistance
R_o = Total resistance of load circuit, $R_o = R_L + R_g$
t_0 = Time of one half cycle
T = Time response constant
V_s = Supply voltage impressed on gate circuit
V_c = Control voltage
ΔV_c = Change of control voltage
V_{c0} = Bias voltage
$\Delta V_{c(n-1)}$ = Total change of half-cyclic average value of control voltage at $(n-1)$th half cycle
ΔV_{c_n} = Total change of half-cyclic average value of control voltage at nth half cycle
W_c = Control winding
W_g = Gate winding
α = Phase lag angle
ϵ = Basis of natural logarithm
λ = Ratio of carrier to signal frequency
τ = Time constant
ω_s = Signal frequency in radians per second

References

1. MAGNETIC AMPLIFIERS (book), H. F. Storm. John Wiley & Sons, Inc., New York, N. Y., 1955.

2. TRANSIENTS IN LINEAR SYSTEMS (book), M. F. Gardner, J. L. Barnes. John Wiley & Sons, Inc., 1947.

3. MAGNETIC AMPLIFIER ANALYSIS AND APPLICATION USING BLOCK-DIAGRAM TECHNIQUES, Richard M. Hubbard. *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 76, Nov. 1957, pp. 578–88.

4. ON FEEDBACK IN MAGNETIC AMPLIFIERS, PART I—SINGLE FEEDBACKS, L. A. Finzi, J. J. Suozzi. *Ibid.*, vol. 77, 1958 (Jan. 1959 section), pp. 1019–31.

◆

Discussion

G. E. Lynn (Westinghouse Electric Corporation, Baltimore, Md.): The author is to be thanked for an interesting discussion of the problem of transfer function representation of magnetic amplifiers. The literature has lacked a comprehensive paper on the subject and Mr. Manteuffel's contribution goes a long way toward filling the void. Several aspects of this subject have, however, been covered in previous papers by various authors[1–3] and it seems fitting that the attention of the interested reader be directed to the earlier work.

A particularly illuminating correlation was established by Mr. Manteuffel in comparing equations 14 and 40 in his paper. In working with an equivalent resistance, R_d, the fact that the time constant cannot be less than $1/2f$ is occasionally lost sight of. The treatment accorded this point in the paper should fix it thoroughly in the reader's mind.

REFERENCES

1. ANALYSIS OF MAGNETIC AMPLIFIERS BY THE USE OF DIFFERENCE EQUATIONS, P. R. Johannessen. *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 73, 1954 (Jan. 1955 section), pp. 700–11.

2. MAGNETIC AMPLIFIER ANALYSIS AND APPLICATION USING BLOCK-DIAGRAM TECHNIQUES, Robert M. Hubbard. *Ibid.*, vol. 76, Nov. 1957, pp. 578–88.

3. A FAST-RESPONSE FULL-WAVE MAGNETIC AMPLIFIER, G. E. Lynn. *Ibid.*, vol. 77, Mar. 1958, pp. 37–41.

Vol.
Pt. 1
196

Magnetization in Tape-Wound Cores

R. C. BARKER
ASSOCIATE MEMBER AIEE

IN RECENT years it has been noted in several phases of the electronics industry that advanced engineering developments follow the developments in related components. As a more extensive knowledge of the components and their operation becomes commonplace among engineers, their application becomes greatly accelerated. An example of this is found in the technology of thermionic devices, where the knowledge of electron dynamics led from the vacuum tube to the cathode-ray tube, the photomultiplier, the traveling-wave tube, and a host of other components. Also, in a few short years we have seen advances in the knowledge of semiconductors lead from the dry-disk rectifier to more advanced semiconductor junction rectifiers, transistors, and other components, which seem to be increasing in number almost monthly.

The same pattern of events characterizes the development of modern magnetic materials. The early technology surrounding soft magnetic materials was developed from a fairly crude understanding of magnetic properties, and almost no knowledge of the basic theory of ferromagnetism. Recent developments in magnetic materials have lead to spectacular advances in their application, resulting in a technology generally called magnetic amplifiers, for several years. But developments have been so rapid, it now appears that magnetic amplifiers will soon represent only a small portion of a larger technological field[1] more conservatively called "nonlinear magnetic devices."

It frequently happens that in the rapid development of such a field, the new applications appear in the literature where they can be easily found, while the more basic descriptions of the components are found scattered throughout the literature. An occasional tutorial paper is sometimes helpful in summarizing the knowledge of the components. Accordingly, the purpose of this paper is to summarize the state of knowledge of soft magnetic materials, specifically as used in commercially available tape-wound cores. The summary is written from the electrical engineering standpoint, and the language is purposely nonmathematical. The intent is to give a qualitative description of the physical and electrical properties of magnetic tapes in terms of simple experiments that can be performed in most laboratories without excessive special equipment. The first section of the paper is basic enough to be applicable to all magnetic materials, whether hard or soft, metallic or nonmetallic. The remainder, however, is directed toward tape-wound, toroidal cores. The alloys discussed are

---

Paper 60-873, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting, Atlantic City, N. J., June 19–24, 1960. Manuscript submitted November 4, 1959; made available for printing June 3, 1960.

R. C. BARKER is with Yale University, New Haven, Conn.

The author wishes to acknowledge the co-operation of his colleague, Dr. Charles B. Wakeman, and of Magnetics, Inc., Butler, Pa. Much of the original work described was done under the sponsorship of the Air Force Flight Control Laboratory, *WCLCX-2*, at Wright Air Development Center.

the 50%/50% nickel-iron, 80%/20% nickel-iron, and 3%/97% silicon-iron alloys, but the general line of reasoning is applicable to many other combinations, even though their properties may be entirely different.

Having obtained a qualitative understanding of the properties of these materials, the reader may wish to investigate the metallurgical or magnetic properties further. It is hoped that the references[1-40] will be found useful for this purpose. The list is not complete by any means, but the items listed include many references to further literature in the fields of specific interest.

Five important sources of information that represent recent advances in magnetism and magnetic materials are the proceedings of conferences that have taken place in the United States and Great Britain in the past several years.[3-6] There are also several textbooks that deal with ferromagnetism,[7-10] and two excellent reviews[11,12] that bring the subject up to date.

Magnetization in Single Crystals

BASIC THEORY

A single crystal is characterized by its uniform lattice structure, disturbed only by occasional lattice imperfections and impurities. Because of the periodic structure of the lattice in the 3-dimensional space occupied by the crystal, many theoretical problems can be formulated by means of group theory in a relatively simple manner.[13] When the lattice is not periodic, it becomes virtually impossible to formulate such problems. Polycrystalline materials fall into this category. They are made up of small crystals. Those varying in size from a fraction of a millimeter are commonly called crystallites. When the size increases to a fraction of an inch or larger they are commonly called grains. A periodic lattice structure exists within the crystallites or grains, but on a larger scale the lattice structure is obviously not periodic.

Thus, the single crystal constitutes the ideal solid in which ferromagnetism occurs, and consequently the introductory discussion of ferromagnetic domains applies to single crystals. Interaction forces leading to ferromagnetism have been calculated and experiments have been performed on some single crystals specially grown for the purpose. The ferromagnetism of polycrystalline materials is treated as a modification of the single-crystal model.

The discussion to follow deals with the ferromagnetism of magnetically soft materials, beginning with the qualitative description of ferromagnetic domains. The atomic theory and energy-band theory are not discussed because they are highly theoretical in nature and require an extensive background in quantum physics. However, it is difficult to reject the topic categorically without making at least one or two comments to summarize the state of knowledge and to give references to the interested reader.

An electron associated with an atom is found to have an intrinsic magnetic moment arising from a hypothetical "spin" about its hypothetical "axis." The directions in which this electron spin can point relative to the electron distribution about the atom are governed by quantum mechanics, and the behavior of this little permanent magnet is not at all analogous to classical magnets. In some materials, it is found that when electron spins on adjacent atoms are parallel, such that their magnetic moments add, the energy of the material is less than if they are antiparallel, such that their magnetic moments cancel. The "forces" that bring this about are called exchange forces; they are electrostatic rather than magnetic in origin, and have no classical analog. The conditions under which the exchange forces are large enough to produce ferromagnetism are found only when the radius of certain electron shells and the interatomic spacing in the crystal lattice are related in just the right way. This explains why it is that only a small number of elements in a restricted portion of the periodic table are ferromagnetic. It also accounts for the fact that alloying and processing, which affect the type and dimensions of a crystal lattice, are important in determining magnetic properties.

The theory of ferromagnetism is far from complete, and important contributions are still found from time to time in the current literature. Thus, in addition to the references listed on the general theory of ferromagnetism, the reader is advised to consult the recent and current literature for the latest developments.[14-18] In the remainder of this paper, the discussion will deal with relatively large-scale models, and atomic scale effects will appear only in the background.

FERROMAGNETIC DOMAINS

The theory of ferromagnetic domains and the molecular field was introduced by Weiss in 1907[11,19] in order to account for two simple and commonly known experimental facts:

1. By applying a very weak field, say 5 ampere-turns/meter (0.063 oersted), it is possible, in a ferromagnetic specimen, to obtain a flux density of approximately 1.5 webers/ square meter (15 kilogausses), a very much larger value than that obtained in air or in paramagnetic substances.
2. It is possible for a bar-shaped specimen of the same type of material to have practically zero external magnetization with zero applied field.

The thermal energy of atoms at room temperature is so large that their random motion would require a strong field to produce alignment of their magnetic moments if acted upon independently. Thus Weiss proposed that there is an internal or molecular field acting to align the elementary magnetic particles. The molecular field was assumed to be the result of the interaction, which we now call exchange interaction, of each elementary magnet with its neighbors, and calculations indicate that it would have to be in the neighborhood of 8×10^8 ampere-turns per meter, or about 10^7 oersteds. This is in the order of 10 times the largest field ever produced in the laboratory. With the assumption that this field is present, it is difficult to imagine that the very small field mentioned could reverse the magnetization of the specimen, or that the specimen could have nearly zero magnetization with zero applied field. Weiss recognized that the molecular field is a point function of space that can vary its direction throughout the volume of the material. Thus, by postulating thin boundaries between saturated regions of the material, Weiss was able to account for the absence of apparent magnetization external to the sample while maintaining completely saturated domains with boundaries placed so that the magnetic flux path closes within the material.

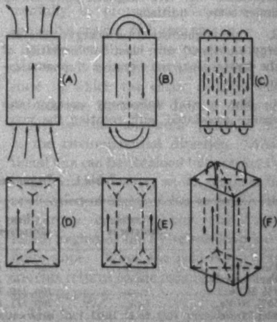


Fig. 1. **Magnetic states of a bar sample**

The state of exactly zero external field was not necessary to Weiss's argument, and is rarely possible experimentally. Fig. 1 shows the manner in which the boundaries between fully saturated regions in a bar-shaped specimen can be placed so as to have a very large external field, a very small external field, and virtually zero external field. The arrows show the directions of saturation of the material, and the dotted lines show the boundaries between the saturated domains. These boundaries are often called, "domain boundaries," "domain walls," or "Bloch walls." In Fig. 1(D), the vertical wall is called a 180-degree wall, and the four short walls extending to the corners of the specimen are called 90-degree walls. The upper domains, which are saturated horizontally, are called domains of closure, because they allow the magnetic flux to form a closed path within the material. Even in the specimen of Fig. 1(D), however, the external field is not exactly zero. There are small local fields in the vicinity of the domain boundaries. Because of these local fields concentrated near domain boundaries, it is possible to place a suspension of very fine magnetic particles on a surface such as that shown and observe the motion and positions of domain boundaries.[11,20]

The sketches shown in Fig. 1 are typical for plane symmetry throughout a thin slab. In thicker specimens, the domains may take on a 3-dimensional pattern such as that shown in Fig. 1(F).

THE ENERGY OF MAGNETIZATION

There are four components of energy involved when a specimen is magnetized. These are:

1. The exchange energy: the difference between the energy of the specimen with antiparallel electron spins and that with parallel spins.

2. The magnetostatic energy: the energy associated with free magnetic poles set up under some conditions.

3. The magnetostriction energy: the energy associated with elastic deformation of the crystal along the direction of magnetization.

4. The crystal anisotropy energy: the energy associated with rotating the magnetization vector with respect to the crystal lattice, away from the preferred or "easy" direction of magnetization.

The first of these components is dependent only on the relative directions of neighboring spins and not on their relation to the crystal lattice. The other three are definitely dependent upon the the direction of magnetization with respect to the lattice, the microstructure within the materials, and the over-all geometry of the specimen. Thus, we say the exchange energy is isotropic, and the other three are anisotropic. The anisotropy energy referred to frequently in the technical literature may include one or more of the latter three components. Common crystal systems are discussed by Barrett,[21] and the crystal systems of common ferromagnetic alloys and their directional properties are treated extensively by Bozorth.[7] A very brief summary of these properties is given in Table I, where the axes of magnetization are given in the order of increasing difficulty of magnetization. It should be noted that the nominally 80% nickel-iron alloy (the Permalloy series) has a very low anisotropy constant, which means that it does not require much more energy to magnetize it in its most difficult direction than it does in its easiest direction.

As a result of our discussion of the four energy components of magnetization, we can now comment on the various magnetic states represented by Fig. 1. In Fig. 1(A), the magnetization is in the same direction throughout the material. The energy of magnetization in this specimen, as in the other three, depends upon whether the bar is cut parallel to the crystal cube edges or otherwise, because of the crystalline anisotropy and magnetostriction constants. For the present, let us suppose that the specimen is cut parallel to the crystal axes. Then the principal reduction in energy that can take place is the reduction of the magnetostatic energy associated with the large external field. In Fig. 1(B), the magnetostatic energy is reduced, but a domain wall has been included. In the domain wall, adjacent atoms do not have parallel spins, as shown in Fig. 2, a model suggested by Kittel.[11] The rapid directional change of the magnetization vector within the wall indicates that anisotropy and magnetostriction energy terms are present. Thus, the reduction in magnetostatic energy in Fig. 1(B) is partially offset by the increase in energy associated with the domain wall within the material. Whether the configuration of Fig. 1(A) or Fig. 1(B) is energetically favorable depends upon whether the magnetostatic energy reduction minus the wall energy gain is positive or negative. If it is positive, Fig. 1(B) is energetically favorable over Fig. 1(A). In the more common iron alloys, the subdivision normally continues as in Fig. 1(C) until a rather large number of domains is formed, perhaps as many as 1,000 per centimeter.

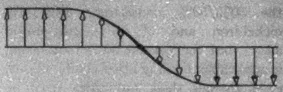


Fig. 2. Vectors representing the change in magnetization through a domain wall

Table I. Properties of Magnetic Alloys

	Crystal Structures	Easy Directions	Crystal Anisotropy Constants in ergs/cm^3	
Pure iron	b.c.c.	*100, 110, 111*	$K_1=42\times10^4$	$K_2=15\times10^4$
Pure nickel	f.c.c.	*111, 110, 100*	$K_1=-34\times10^4$	$K_2=53\times10^4$
3% Si-Fe	b.c.c.	*100, 110, 111*	$K_1=28\times10^4$	$K_2=10\times10^4$
50% Ni-Fe	f.c.c.	*100, 111, 110*	$K_1=3\times10^4$	$K_2=-18\times10^4$
80% Ni-Fe	f.c.c.	substantially isotropic	$K_1=-0.2\times10^4$	$K_2=-2\times10^4$

In many materials, such as the alloys listed in Table I, there are three preferred directions of magnetization because of the symmetry of the cubic crystal systems. In such cases, it may be found that the magnetostatic energy can be reduced nearly to zero without the formation of a large number of domains, by the introduction of domains of closure as shown in Fig. 1(D). The four magnetization vectors are all in equally preferred directions. However, since the magnetostriction results in an elongation, or contraction, in the direction of the magnetization vector, there is an elastic deformation in the specimen that represents energy. Whether the closure domains form, and if so the number that form, depends upon the relative magnitudes of the energies represented and the method by which the material is magnetized. The history of magnetization is important because there are often quasi minima of energy that do not correspond to the absolute minimum.

This brief and qualitative description suggests that the magnetization process takes place by the growth of domains saturated in one direction at the expense of those saturated in another direction. The magnetization of modern soft magnetic materials is unmistakably a process of this kind, even in polycrystalline specimens. However, even when the material consists of just one fully saturated domain, the magnetization vector, which is parallel to the preferred crystallographic direction,

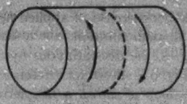


Fig. 3. Cylindrical shell with zero total flux

may not be parallel to the applied field. Thus, further flux may be obtained in this direction by rotating the magnetization vector into alignment with the applied field. This is done at fields higher than those required to propagate domain walls because of the anisotropy energy characteristic of the particular crystal. Domain rotation will be discussed further in a later section.

MAGNETIZATION IN THE THIN CYLINDRICAL SHELL

The toroidal core is a structure that has essentially no air gap. Therefore, the magnetostatic energy does not play as important a part in the magnetic properties of the specimen as it does in the rectangular bar of Fig. 1. Consequently, it is to our advantage to develop the concepts of magnetization by domains by means of a closed configuration. It is convenient to use the cylindrical shell as the starting point. The cylindrical shell is shown in Fig. 3. It is assumed for this example that the shell is a single crystal with its cubic axes directed circumferentially, axially, and normal to the surface. A domain wall is shown dividing two regions of complete saturation. According to our previous qualitative argument, there is a certain amount of energy associated with this domain wall, and this should not depend upon its location in the shell. Thus, it should require zero energy to move the domain wall from one edge of the shell to the other, i.e., to reverse its magnetization. However, when the shell is completely saturated, the domain wall is absent entirely. To reverse the magnetization from one completely saturated direction to the other requires that a field be applied sufficiently large to establish the domain of reverse magnetization. This field is commonly called the "starting field" or the "nucleating field." Once the nucleus of reverse magnetization has been established, it requires a vanishingly small magnetizing force to propagate it through the shell to the other edge.

The hysteresis loop of the ideal cylindrical shell can be obtained by considering it to be fully saturated, as shown in Fig. 4(A), and applying the field necessary to establish a nucleus of reverse magnetization and propagate it through the shell, leaving it fully saturated in the opposite direction. Fig. 4 shows the hysteresis loop implied by the ideal shell. The total flux ϕ is plotted against the magnetomotive force F. The regions corresponding to the four states of magnetization shown in the figure are indicated by arrows. It is assumed that the magnetomotive force is increased in the negative direction by the experimenter until the domain wall is established, and is then decreased rapidly to zero. The magnetomotive force is then increased to a vanishingly small value, resulting in the motion of the domain wall from the position shown in Fig. 4(B) to the position shown in Fig. 4(C). Further increase in the applied field moves the domain wall out of the shell and leaves it fully saturated in the negative direction as shown in Fig. 4(D). The other half of the hysteresis loop is shown dotted, and is the image of the first half.

In the above described process of flux reversal, it was assumed for the sake of simplicity that the domain of reverse magnetization was nucleated at the left edge of the cylinder. One might think that the return to positive saturation would be accompanied by reverse motion of the domain wall, i.e., from right to left. If there are any physical factors that cause the first reversal to begin at the left edge of the cylinder, these same factors would cause the second reversal to begin also at the left edge. In normal crystals, domains of reverse magnetization would become established at impurities or imperfections near the surfaces, and propagate outward into the cylinder. In the assumed perfect crystal, these centers of flux reversal do not exist, and the domains of reverse magnetization would probably appear at both edges simultaneously at a somewhat higher nucleating field.

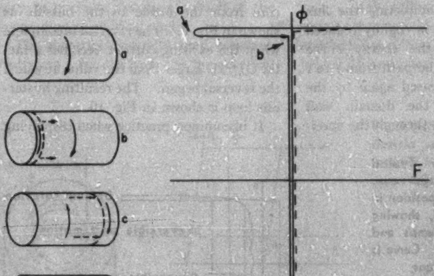


Fig. 4. Ideal hysteresis loop of cylindrical shell, correlating points on the hysteresis loop with domain-wall position

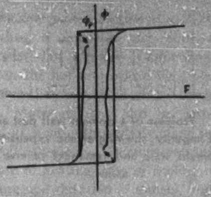


Fig. 5. Hysteresis loop of a single crystal showing domain-nucleation effect

The foregoing is a first approximation to the mechanism of magnetization in a single crystal. The description is qualitative and not rigorous. Any further detail at this stage of the argument, however, would be superfluous because, even in a single crystal, this degree of idealization cannot be realized. Even if the thin shell could be grown as a single crystal, there are two factors that would have to be considered. The first of these is that the crystal orientation would not be exactly directed along the circumference, and secondly, it is not possible experimentally to propagate a domain wall through a crystal with zero energy.

It the event that the crystal orientation is not exactly circumferential, the state in which the entire cylinder constitutes one saturated domain does not result in the maximum possible flux in the circumferential direction. Additional flux can be obtained by rotating the magnetization vector, which requires the application of higher fields to supply the anisotropy energy. Thus, the "saturated" region of the hysteresis loop has a finite slope. When the field is removed, the flux returns to the remanent point, ϕ_r, shown in Fig. 5.

The fact that it is not possible experimentally to move a domain wall with

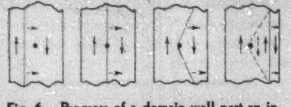


Fig. 6. Progress of a domain wall past an included impurity, showing elastic expansion of the domain wall and remanent reverse domains

zero field even in a single crystal means that there is some "coercive force" required: the flux-current loop has a finite width. The theory of coercive force is still not understood in quantitative detail, but qualitatively, it is believed with good experimental authority that the existence of coercive force is a direct evidence of crystal lattice imperfections and impurities. Experiment shows[7] that when the field is increased until the flux in the core begins to change, this value of field may be sufficient to produce the entire flux reversal. If after the flux begins to change the field is reduced, the flux is prevented from changing further. The field may then be increased again from a small value, and the flux is found to reverse in steps or jumps. The resulting flux-current loop is shown in Fig. 5. It should be noted in Fig. 5 that the field required to propagate the domain through the material is less than the nucleation field. As indicated by Kittel and Bozorth in the foregoing references, these experiments have been made on single crystals[22] of somewhat different dimensions than the cylindrical shell, and polycrystalline materials that have been annealed in the presence of a strong magnetic field applied in the direction of magnetization. Very nearly the same results can be found in commercial-grade tape-wound cores, as will be shown in a later section.

The dependence of coercive force upon impurities and lattice imperfections is worthy of additional comment, since it accounts qualitatively for the "initial permeability" of materials and the well-known Barkhausen discontinuities. The crystal lattice has in it some imperfections such as slip planes, dislocations, missing planes of atoms over short distances, and impurities such as precipitated concentrations of foreign atoms (called inclusions) and some foreign atoms substituted or placed interstitially in the crystal lattice. The percentage of impurity atoms is small and the percentage of volume occupied by imperfections and impurities is small; nevertheless, they are present. Thus the domain wall that moves through the material does not move smoothly, nor does it always have the same length. When a domain wall is pulled over an impurity, or imperfection, it tends to stick to the impurity. As the motion of the domain wall continues, the wall elongates and finally snaps free. The motion of the domain wall is analogous to a taut rubber band being pulled over a rough piece of sandpaper, sections of the band snapping from one piece of grit to another as it moves. Each time a section of the domain snaps forward, the volume through which it sweeps reverses its magnetization. It is easily imagined that these volumes range widely in size. The resulting random discontinuities in flux reversal are called Barkhausen discontinuities, and the noise detected in the material is called Barkhausen noise. The situation described is shown in Fig. 6.

If we could plot the energy of the domain wall as it moves through the shell, a curve such as that shown in Fig. 7 would result. The energy rises in certain areas as the domain wall elongates while passing an impurity, and decreases as it snaps free. If the domain wall is located near an energy minimum, for example at *a* in Fig. 7, changing the applied field by a small amount carries the energy toward *b* or *c*. This is an elastic motion corresponding to an elongation of the domain wall. If the field is removed before the energy either at *b* or at *c* is reached, the wall gives up its energy and returns to *a*. The corresponding small reversible change in magnetic flux with applied field results in the so-called initial permeability. If the field is continually increased, the energy curve follows the path shown with irreversible jumps from *c* to *d*, *e* to *f*, and finally from *g* to *h* completing the flux reversal. If the field is rapidly reduced when jumps occur, the energy curve follows, for example, the path from *f* to *i* until the field is increased again by the experimenter. After the domain wall has moved completely through the specimen, further increase of applied field results in domain rotation.

In Fig. 8, the discontinuous hysteresis loop of Fig. 5 is shown for reference. Inside this loop is shown a magnetization path, a portion of a "minor loop," representing the three regions mentioned above: reversible domain boundary motion, irreversible boundary motion, and domain rotation. This classification of regions is only approximate, however, and applies only to relatively slow changes.

CONCENTRIC CYLINDRICAL SHELLS

For the present purpose, it is convenient to construct a toroid of more reasonable physical dimensions by imagining a large number of our cylindrical shells assembled tightly and concentrically together. It will then be necessary to modify our results in accordance with the particular crystal structure in mind, and the fabrication details of the toroid.

When a toroid is wound with N turns and a current of I amperes is applied, the total magnetomotive force applied to each circumferential magnetic path is the same, namely, NI ampere-turns. But the circumferential magnetic paths are longer at the outer diameter of the toroid than they are at the inner diameter. Thus, the magnetic intensity, NI/π ID (inside diameter) is larger at the inner diameter than it is at the outer diameter, NI/π OD (outside diameter) in ampere-turns per meter. Let us consider first the innermost cylindrical shell, and for the moment consider that the concentric shells magnetize independently. When the magnetomotive force reaches the value corresponding to the nucleating field for the inner shell, its magnetization will reverse. As the current is increased further, the flux reverses in cylinders of increasing diameter. Thus a cylindrically shaped domain wall is formed that increases its diameter, thus reversing the flux in the core from the inside to the outside as shown in Fig. 9. The reversal is complete when the exciting current becomes a factor OD/ID larger than the value at which the reversal began. The resulting hysteresis loop is shown in Fig. 10.

It is common practice when displaying

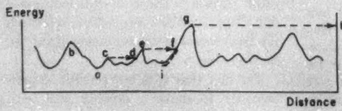
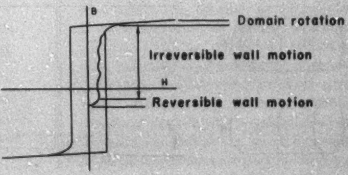


Fig. 7 (left). Typical curve of energy versus domain-wall position in single crystal, showing Barkhausen jumps and quasi minima. Curve is not unique

Fig. 8 (right). Contributions to the magnetization process



hysteresis loops to plot flux density **B** against magnetic intensity **H**. The flux density is obtained by dividing the measured flux by the effective cross-sectional area of the core. There is no ambiguity in this process of normalizing, but there may be some error because one does not usually know the exact per cent of the core area occupied by the iron. In normalizing the magnetomotive force to obtain magnetic intensity, it is usual to use the mean magnetic path length. Here, there is some possibility of confusion. The normalizing procedure makes it possible to compare the coercive forces of different materials without having to obtain identical specimens. However, it should be recognized that different portions of the hysteresis loop correspond to flux reversals in different portions of the core. Thus, when one sees magnetic intensity **H** on a hysteresis loop, he should remember that this is the value of magnetic intensity only at the reference magnetic path length. The value of **H** for which the core saturates, in other words, is the magnetic intensity at the center of the core and not at the outer portion of the core, where the flux is changing.

If now, we consider the effect of adjacent concentric shells on each other, we find that after we have exceeded the nucleating field at the inner diameter of the core, there is more than enough magnetomotive force present to cause the nucleated domain to move some distance out into the toroid. This is similar to the spontaneous flux reversal found in the cylindrical shell. But now the domain is approximately cylindrical in shape and is propagated outward into the toroid and grows in size and therefore energy. After it has moved a short distance, a further increase in magnetomotive force is required to produce further flux reversal. The resulting hysteresis loop is shown in Fig. 10. In this case, it is no longer necessary to nucleate domains in each of the succeeding cylindrical shells.

The materials actually used commercially are a far cry from the hypothetical single crystals assumed in the foregoing descriptions. The concepts and simple models developed through these assumptions, however, have a close relationship to the behavior of commercial cores at power frequencies. We shall demonstrate this shortly. But first it is necessary to describe the actual crystalline nature and fabrication of these cores.

Microstructure of Polycrystalline Materials

Commercially available tape-wound cores are fabricated from materials melted under carefully controlled conditions, occasionally in a vacuum where exceptionally high purity is required, but usually under a protective slag. The melt is then cast into ingots and hot-rolled down to a thickness in the neighborhood of 0.1 inch. From this point the material is cold-rolled and annealed in stages down to the final thickness, where it is sheared to the desired width. The material is then sprayed with a magnesium oxide suspension and wound spirally into tape cores which are given a final stress-relief anneal, usually in an atmosphere of dry hydrogen.[7]

With a given alloy, the details of the cold rolling and annealing schedules are crucial in bringing about the magnetic properties desired in the finished product. The details of the rolling and annealing schedules depend not only on the alloy being processed, but also on the finished gage desired. Tape-wound cores are commonly available in 1/8, 1/4, 1/2, 1, 2, 4, and 6 thousands of an inch (mils) thick. Thickness is also given sometimes in microns, where a micron is equal to 10^{-6} meter. A convenient figure to remember is that 25 microns correspond to approximately 1 mil. Thus, for a given alloy, the rolling and annealing schedules may differ greatly from the thickest to the thinnest gage. In the rolling part of the final processing, the variables are the thickness at which hot rolling stops and cold rolling begins, the number of cold-rolling stages, and the degree of reduction at each stage. Annealing is normally not done after each pass, but is more commonly done between the hot and cold rolling phases, and after the material reaches its finished gage. Materials may also be strip annealed at some intermediate stage in the cold-rolling process. In any given anneal, the rate of heating, rate of cooling, temperature and duration of anneal, and the annealing atmosphere are all important variables. It is frequently found that different lots of the same alloy processed to the same gage require different annealing schedules in order to produce equivalent results. With the large number of variables mentioned, it is not surprising that the techniques required to achieve the desired magnetic properties are largely determined empirically. It is also quite likely that processing schedules that differ markedly in detail can be made to produce the same equivalent results.

It is perhaps appropriate to mention at this point that the theory of annealing and the detailed mechanisms by which cold working and annealing determine the microstructure of metals are not understood. However, many of the effects are well understood qualitatively. The factors of importance in magnetic alloys are grain size, relative grain orientation, the character of grain boundaries, the orientation of the lattice in each grain with respect to the plane of rolling, and impurity content. All of these factors are dependent upon cold working and annealing, and it is these factors that one tries to control in producing high quality magnetic materials. Advances in the theory of these processes are reported in the current literature from time to time, and are discussed in detail by Bozorth.[7]

Unfortunately, there is no single reference in the literature, to the author's knowledge, where all the inportant properties of commercial-grade alloys are listed. Most of the information given in this

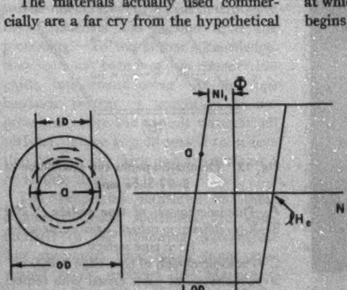


Fig. 9 (left). Radial expansion of domain wall and corresponding increase in magnetizing force required to propagate it

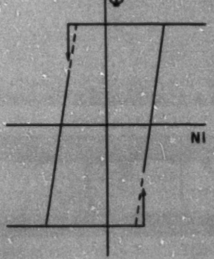


Fig. 10 (right). Hysteresis loop of toroidal crystal showing spontaneous flux reversal following nucleation

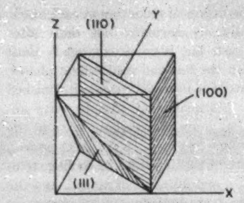


Fig. 11. Scheme for designating planes in a cubic crystal

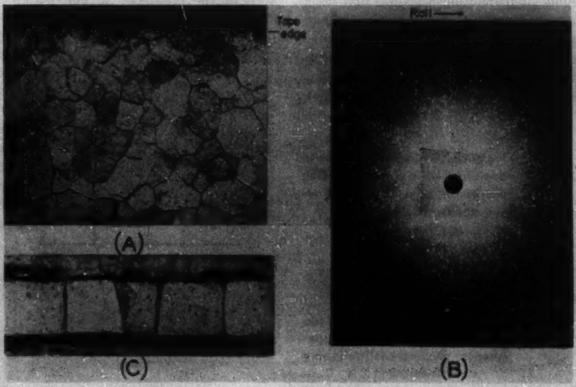
paper comes from references scattered throughout the literature, from private conversations, and from measurements made at Yale University in the course of past research. Further information on the properties of polycrystalline materials can be found in the book by Bozorth,[7] the current literature,[2-6] a pair of excellent descriptive articles by Randall and Scholefield,[23,24] and the doctoral thesis of Wakeman.[25] Many of Wakeman's results are quoted in this paper.

In the present discussion, we are mainly concerned with the nickel-iron alloys, and 3-97 silicon-iron. It is believed that the effects of cold rolling and annealing are qualitatively the same in these materials. But the discussion applies principally to the nickel-iron alloys. The effects of drastic cold reduction and annealing in the thin gages seem to be most spectacular for these alloys. Consequently, they have been studied in more detail.

Hot rolling is used principally to take the materials down close to their finish gages. The process usually leaves some plastic strains which are removed by annealing at a high temperature, 1,100 C (degrees centigrade) or higher. The process of annealing gives the atomic lattice a large amount of thermal or vibrational energy. This allows some lattice deformations to heal, thus relieving plastic strains. It also tends to break up precipitated impurity inclusions and encourages the impurity atoms, which often are too dissimilar to the principal alloy to substitute in the crystal lattice, to diffuse to the surface where they evaporate or combine chemically with the atmosphere in the furnace. It is also possible that atoms of the annealing atmosphere diffuse into the metal and combine chemically with the immobile impurity atoms, thus making them mobile. The combination may then diffuse out of the material. In any event, all impurities are not removed. A further function of annealing is that it promotes the growth of large regions or grains which are single crystals. A qualitative explanation for this effect is that sufficient thermal energy is given to the atoms that the identity of the lattice structure is lost, or partially lost. In cooling, impurities or intact crystallites act as seeds, and large grains grow with these as centers. This is the situation that is found after the anneal subsequent to hot rolling. The large grains attest to a high degree of purity and relatively low internal strains. These grains may be in the order of half an inch or larger, and can frequently be seen with the naked eye by reflected or polarized light on an oxidized or etched surface.

After hot rolling to about 0.1 inch, and annealing at a fairly high temperature—1,100 to 1,200 C for the nickel-iron alloys—the material is ready for cold working. In thin gages, the nickel-iron alloys are frequently reduced by 98% or more. It is generally found that the beneficial effects of cold rolling are enhanced by making the reduction in many passes. The first stage of cold reduction shatters the large grains leaving very small crystallites, even though the material has not been heavily worked. After as little as 10% cold reduction, the large grain boundaries are still visible in photomicrographs. But X-ray reflection patterns show that the grains are shattered into small undiscernible crystallites.[23] Subsequent heat treatment, however, usually allows the regrowth of these grains. Thus the cold working is generally continued without intermediate annealing. Above 75% cold reduction, the finely divided grains become noticeably smeared out in the direction of rolling by the mechanical action of the rolls. This severe cold working also brings about the orientation of the crystallographic axes of the crystallites. This effect becomes more pronounced with further cold reduction, until at 95% reduction or greater, the crystallites have their crystallographic axes oriented to within 5 degrees or less of the direction of rolling, and the material has a distinctly shredded and fibrous appearance.[23] In the materials with a high degree of cold reduction, annealing at high temperatures usually induces some regrowth of crystal grains. The size and orientation of these grains depend upon the degree of cold reduction and the final tape thickness for any given alloy.

This description is rather general and qualitative. It is most descriptive of the nickel-iron alloys, where the degree of orientation and the effect on the magnetic properties is most pronounced. In the 50-50 nickel-iron alloy the grains are oriented to a high degree, with the cube edges directed along the direction of rolling. The cube edge is also the direction of easy magnetization. In the 80-20 nickel-iron alloy group, the crystallites become oriented with their body diagonals



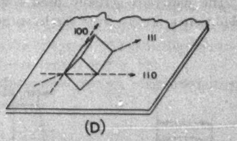
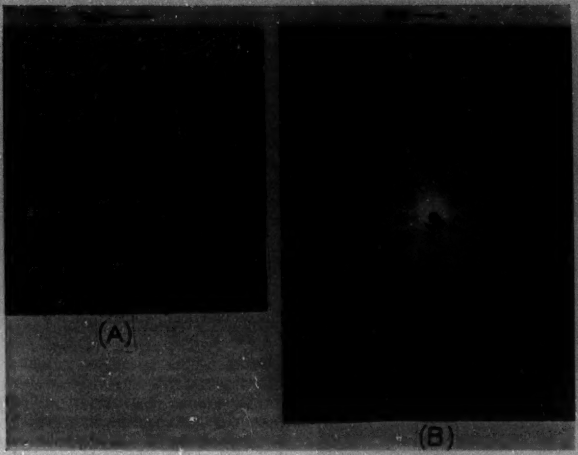


Fig. 12. Photomicrographs and orientation of 3–97 Si-Fe tape

A—Photomicrograph of tape surface ×200
B—X-ray diffraction pattern, beam perpendicular to tape surface
C—Photomicrograph of tape edge ×800
D—Orientation of cubic crystal with respect to tape



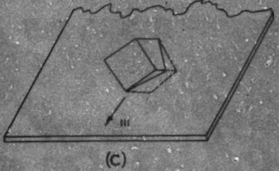


Fig. 13. Photomicrographs and orientation of 4-79-17 Mo-Ni-Fe tape

A—Photomicrograph of tape surface ×800
B—X-ray transmission pattern, beam perpendicular to tape surface
C—Orientation of cubic crystal with respect to tape

along the direction of rolling, and become oriented at much thinner gages than the 50-50 alloy. The cube edge, however, is still the direction of easy magnetization. Thus the magnetization in the tape is not in the easy direction. The 3-97 silicon-iron achieves some degree of orientation, but the degree is orders of magnitude below that of the 50-50 nickel-iron alloy.

In the heavier gages, of approximately 4 mils and up, the surface effects may become significant. The relatively small rolls used in the Sendzimer mill,[7] for example, are in the order of an inch in diameter, and show definite tendencies for more severe cold working at the surfaces than is achieved at the center of the tape.[28] Thus the grains in the center of the material are often 10 to 100 times the size of the grains at the surface. This results in a stratification of magnetic properties within the material. These surface effects seem to appear in relatively thick materials and depend upon the processing. To the author's knowledge, they have not been found in commercial-grade, tape-wound cores. As the tape becomes thinner, a condition is approached where the tape is virtually all surface. There is good reason to believe that tapes 1 mil thick or less are about 1 crystallite thick.

The properties measured in three commercial-grade specimens are summarized in the following paragraphs. The specimens were taken from tape-wound cores supplied by Magnetics, Inc., in 1954. The preparation of the specimens is described by Wakeman,[26] who carried out the experimental work. For convenience, the scheme for denoting crystal planes is shown for the cubic crystal in Fig. 11. The crystal axes, or directions, are the normals to these planes. Only three of many possible planes are shown.[21] The cubic crystal is the one most frequently encountered in magnetic materials, but not exclusively. Many cobalt alloys, for example, are hexagonal.

3-97 Silicon-iron: Photomicrographs of 1-mil silicon-iron are shown in Fig. 12. Samples are of material marketed under the name of Magnesil. Fig. 12(C) shows the X-ray diffraction pattern for a 2-mil sample. The spotted nature of the diffraction pattern indicates a granular structure as shown also in the photomicrographs. Note in Fig. 12(B) that the grains go completely through the tape. The "spoked" nature of the diffraction pattern, while not as striking as the pattern for Orthonol, indicates that there is a definite tendency to grain orientation. Fig. 12(A) shows that the largest number of grains are between 1 and 3 mils across. Grains in thicker tapes may be as large as half an inch or larger across. In the ideally oriented material, the crystal axes lie as shown in Fig. 12(D) with respect to the tape surfaces. The earliest work directed toward producing oriented magnetic materials was done with silicon-iron, and was reported by Goss.[27]

4-79-17 Molybdenum-nickel-iron: Samples were taken from commercial cores marketed under the name of HyMu-80 Fig. 13(A) shows a photomicrograph of the surface of a 1-mil sample. It will be noted in this photograph that the process of "twinning," which is the first step in crystal fracturing, is not yet complete.[21] This suggests that further cold reduction would result in a more finely divided and oriented structure. The X-ray diffraction pattern of Fig. 13(B) shows the Debye rings characteristic of a randomly oriented crystal structure. The material is generally reduced about 50% in the final roll, which agrees with the relationship between degree of twinning and final roll reported by Randall and Scholefield.[22] Fig. 13(C) shows the orientation of the crystallites found in the highly oriented specimens. The crystallographic orientation is along the *111* axis. The easy axis of magnetization is the *100* axis.

50-50 Nickel-iron: Samples were taken from material marketed under the trade name of Orthonol. Photomicrographs of samples of 1-, 2-, and 4-mil tape show no signs of grain structure in the material. Further acid etching of the surface shows "etch pits," which when analyzed show that the material has a cubic texture, with its 100 directions perpendicular and parallel to the direction of rolling, and normal to the surface. The photomicrograph of the etched surface is shown in Fig. 14(A). The analysis of the etch pits is given by Wakeman.[26] To corroborate these results, an X-ray transmission pattern was made with the beam normal to the surface of the tape. The tape was 1 mil thick. The cubic structure of the material would predict the fourfold symmetry shown in the transmission pattern. But the pattern, Fig. 14(B), also shows eightfold symmetry. Thus the *100* axis of the crystal may just as well be at 45 degrees with respect to the rolling direction and in the rolling plane. This am-

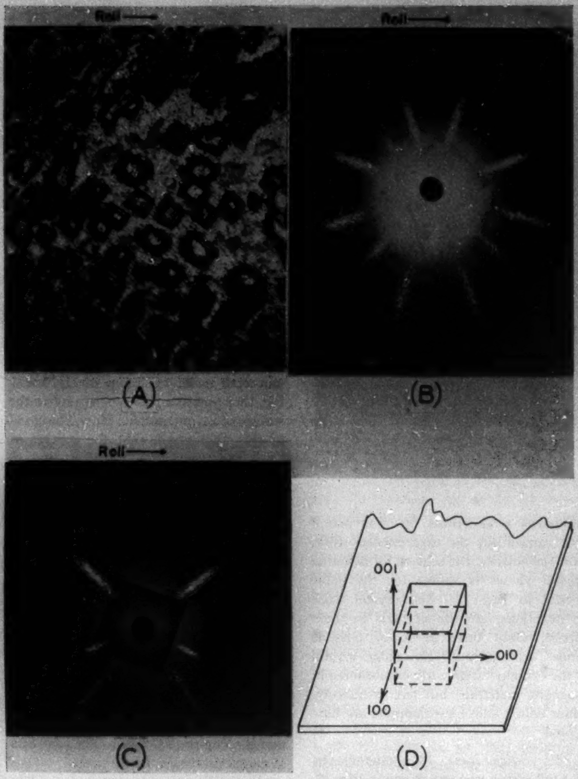


Fig. 14. Photomicrographs and orientation of 50–50 Ni-Fe tape

A—Etch pits ×500. B—X-ray transmission pattern, beam perpendicular to tape surface. C—X-ray transmission pattern, beam 45° to tape surface. D—Orientation of cubic crystal with respect to tape

biguity is removed by tilting the beam 45 degrees with respect to the surface. The resulting twofold symmetry Fig. 14(C), indicates that the *100* axis is indeed parallel to the direction of rolling. The orientation of the cube with respect to the tape surface is shown in Fig. 14(B). If the sample had been a single crystal, the X-ray pattern would have been a series of regularly spaced spots instead of the diffuse spokes shown. This indicates that the crystal sizes are small enough so that a large number lie, within the 1-millimeter X-ray beam used. If, for example, it is assumed that the crystallites are about 1/100 the diameter of the X-ray beam, or 10 microns, the tape is two to three crystallites thick. Also, as in indication of relative sizes, the dimension 10 microns corresponds to about 30,000 atomic layers. The angular spread of the spokes in Fig. 14(B) indicates that the orientation is such that most of the crystallites lie within 3 degrees.

A commercial grade of 50-50 nickel-iron alloy less oriented than Orthonol and its equivalents is also available. The material is marketed under the trade name of 48 Alloy, for example, by Magnetics, Inc., and this is also available from a number of other manufacturers, as are the other alloys discussed. No photomicrographs or X-ray patterns are presently available on this material. It is likely, however, that the same melt is used by a given manufacturer for both this grade of material and the highly oriented grade corresponding to Orthonol. It is probable that tests during the processing determine whether the material will be one or the other, and the processing is then completed accordingly. It is believed that in the 48 Alloy strip, there are some regions with fairly large grains and some regions of finely divided crystallites of fairly high orientation. These regions are probably a fraction of an inch in size, and are of varying shape.

Magnetization in Polycrystalline Tape-Wound Cores

DOMAIN WALLS IN POLYCRYSTALLINE TAPES

We have dealt extensively with the magnetizing process in single crystals even through we do not expect to find them in commercial-grade core materials. It is possible in this manner to discuss the important first-order mechanisms and develop and understanding of the magnetization process without having to complicate the description unnecessarily. Having done that, our task is then to modify the descriptions to take into account the polycrystalline nature of the materials and the fact that the toroid is fabricated by winding a tape spirally, and is not a solid cylindrical section. We shall first describe the effect of polycrystalline structure on domain nucleation, wall motion, and rotation in the magnetic tape. Then we shall describe a model of flux reversal in tape-wound cores. Finally we shall illustrate the model by experiment.

Powder patterns have been made which show the domain walls at the surfaces of polycrystalline materials with fairly large grains. The patterns show that if the grains are sufficiently well oriented, the walls are continuous through the grain boundaries.[11] The situation is aproximately as sketched in Fig. 15. The wall is not straight and therefore is longer and contains more energy than it would in a single crystal. Further, its length continually changes as it moves. Because of the anisotropy energy, the domain walls tend to follow the directions of easy magnetization within the grains. If there were no anisotropy energy required, the domain walls would pass through neighboring crystallites in a straight line without regard for changes in crystal orientation. The shape of the toroidal core, however, helps to keep the flux aligned with the wrapping direction of the tape. If the flux vector were at an angle to the wrapping direction, large magnetostatic fields about the core would be present,

Vol.
Pt. 1
196

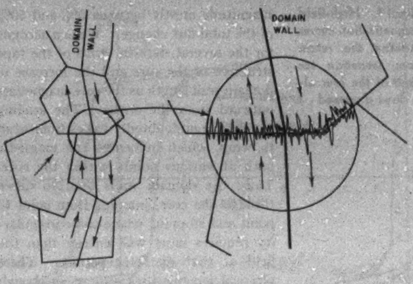


Fig. 15. Continuous domain wall in granular material. Enlarged section shows reverse domains along grain boundaries

requiring a large amount of energy. This gives rise to an effective "shape" anisotropy in the toroidal core.

Breaking the wall at the grain boundary would require that free poles be formed or that complicated closure domains be formed. When the grains are nearly oriented, continuous domain walls require much less energy than is required when the grains have a true random orientation. Thus, the cores with a grain-oriented structure would be expected to resemble closely the single crystal. The grains shown in Fig. 15 are partly oriented.

Note in Fig. 15 that there are a large number of irregular domains of reverse magnetization that cling to the grain boundary. These will persist, changing somewhat in detail, while the domain wall moves through the material. They will be largely but not completely removed with the application of larger field. Some of these domains remain when the material is "completely" saturated, and become nuclei of flux reversal when the field is reversed. Goodenough[28] has extensively treated the subject of nucleating centers and the energy of flux reversal in polycrystalline materials.

In the highly oriented materials such as Orthonol, the crystallites are small and the domain wall passing continuously through large regions is nearly straight. Both from the nature of their crystal structure and measured core properties, it would seem that these materials can be treated as single crystals for most practical purposes. In materials with moderate grain size, as pointed out in the foregoing, it is usually energetically favorable for the domain walls to be continuous over fairly long ranges. When the grains become less oriented, there is a tendency for the long range domain walls to break up into shorter sections. As a result, there is a corresponding tendency for closure domains to form near grain boundaries in order to reduce the number of free poles formed. The resulting domain structure is very complicated. Certainly, there are many more domains of reverse magnetization remaining when the material reaches saturation.

Domain Rotation in Polycrystalline Tapes

When each miniature grain is fully saturated by the application of a field, the direction of its magnetization lies very close to its neighbors and very close to the direction of rolling along the tape. Thus when the domain walls have been moved out of the material, the flux density in the circumferential direction is very nearly as large as can be produced by that volume of material, and is only slightly increased by domain rotation. The contribution of domain rotation to the circumferential flux density may be explained most easily in terms of a partially oriented material. Fig. 16 shows a strip of material of this kind: (A) shows the domain wall moving from left to right in the material, and (B) shows the complete magnetization of each grain, where the magnetization vectors are not closely aligned. The application of larger fields rotates the vectors of magnetization with respect to the crystal axes of the grains, as shown in (C). Larger fields are required because of the anisotropy energy implied in the process of domain rotation. This portion of the magnetization curve is mostly reversible, since the magnetization vectors return to their easy directions when the field is removed. The maximum flux density obtained when all of the magnetization vectors are aligned is called the saturation induction. That remaining when the field is removed is called the residual or remanent induction. The ratio of the two is clearly a measure of the degree of crystal orientation in the direction of rolling.

Domain Nucleation and Flux Reversal in Toroids

The domain nucleation process in polycrystalline tape-wound cores is somewhat different from that found in single crystals. A qualitative explanation, however, follows quite logically from a knowledge of the crystal structure and method of fabrication. As mentioned previously, the exact manner in which the domain walls move in a tape-wound core is not known. It is known, however, that the cores have magnetic characteristics very similar to those found in single crystals. Thus, it is reasonable to suppose that the co-operative phenomena characteristic of the single crystal are also characteristic of the tape-wound core, and that the magnetization by domains proceeds in a similar manner. It is expected that the flux reversal proceeds from the inner diameter outward, and that there may be important magnetic coupling between adjacent layers of tape through the interlaminar space.

Experimentally, it is found that the flux in a tape-wound core does not change discontinuously upon exceeding the starting field, as indicated for the single crystal in Fig. 5. In that example, the flux reversal was attributed to the establishment of a single cylindrical domain wall that propagated outward into the toroid. Even if the cylindrical shape is not a correct assumption, domain walls in the form of some similar surface or surfaces would yield the same general results. But in the tape-wound core, the material is polycrystalline, and the inner surface of the core contains a tape end, and some air gap, which require modification of the reasoning used for the single crystal.

The flux reversal process in polycrystalline tape, discussed in the previous section, leaves some domains of reverse magnetization distributed through the tape at grain boundaries, impurities, and imperfections. These small residual domains act as centers of reverse mag-

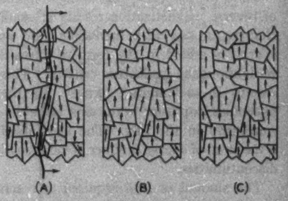


Fig. 16. Flux reversal in granular material, showing

A—Motion of continuous domain wall
B—State of magnetization when domain wall has passed to the right edge of the tape
C—State of magnetization after individual domains have been rotated into alignment

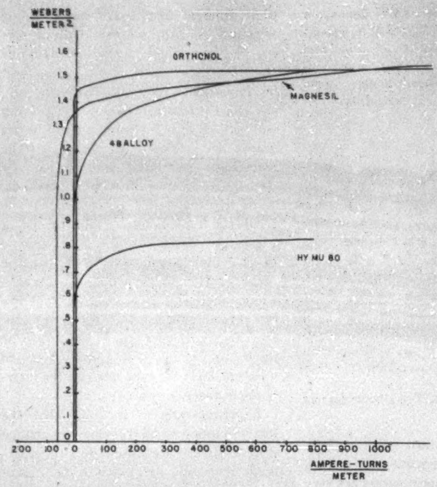


Fig. 17. High-field magnetization curves showing the return from saturation for each of the four alloys discussed

netization when the field is reversed. The type and purity of the alloy determine their distribution.[28,31] The small and isolated regions of reverse magnetization grow independently, resulting in a large amount of domain wall energy for the small amount of flux change. After a certain amount of domain growth, the areas of reverse magnetization begin to overlap. As domain walls begin to combine and shorten it is found that a rather large change in flux can take place without further increase in domain-wall energy. There usually occurs at this point a sudden change in flux with no change in field as the domain walls consolidate and approach some simple surface. The exact shape that the walls take in a tape-wound core is not known.

Thus, the process of nucleating domains of reverse magnetization results in small changes in flux with increasing field as energy is added in the form of domain walls about the nucleating centers. This results in a rounded "knee" on the hysteresis loop prior to the first large discontinuity in the flux. The knee itself is made up of a large number of very small discontinuities.

The amount of field required to reach the point of large discontinuous flux reversals depends upon the number of nucleating centers remaining in the tape after the saturating current has been removed. In Fig. 15, for example, the number of small reverse domains that are present depends upon the magnitude of the saturating field applied before the field is again reversed. The number and size of these domains both reduce with increasing field, and are removed almost entirely for values of field 3 to 10 times the coercive force, depending upon the material. As the field is reduced, rotation of the magnetization vectors back to the easy directions within the grains may leave free magnetic poles at the grain boundaries and precipitates. The total magnetic energy in some cases may be reduced by the formation of the small reverse domains shown in Fig. 15, which tend to alter favorably the distribution of free poles. The distribution of these reverse domains depends upon the material, of course, and the number that are removed at the peak value of the saturating field.

From the above reasoning, it is expected that the field required to nucleate flux reversals at the inner surface of a tape-wound core depends upon the degree of prior saturation of the core. This will be demonstrated by experiment. The amount of flux change corresponding to the process of establishing and expanding domain walls is about 3% of the total flux in a typical Orthonol core. It is somewhat larger in the other alloys.

When the field has reached the value necessary to combine domain walls and produce large discontinuities in flux, the domain walls spread from the inner portion to the outer portion of the core. As mentioned previously, the exact manner of domain-wall growth is not presently known. But the flux reversal takes place in large discontinuous steps, ranging in magnitude mostly between 5% and 50% of the total flux change. We can account for the several steps in terms of the tape structure of the core and the increase in domain-wall length as the core magnetizes outward from the center. The spacing of the tape undoubtedly varies from wrap to wrap, which influences the magnetic path at various points within the core. Thus, the domain walls tend to move through the core, jumping from point to point and stopping where some irregularity requires more wall energy than the field at that diameter permits. These physical irregularities appear as though they were large-scale Barkhausen discontinuities.

EXPERIMENTS

The fundamental character of flux reversal in polycrystalline core materials can be verified experimentally by means of some fairly simple experiments. Some of these are described below with comments that should help the experimenter to reproduce the results for himself. First, we shall show that the measured properties of magnetic tapes described earlier are those that we would expect from their crystalline makeup. Secondly, we shall examine the nucleation and propagation of domain walls in commercial-grade toroids. Finally, an experiment is described in which it is shown that the magnetization of a core proceeds from the inner diameter of the core outward.

Confirmation of the degree of orientation of the various tapes can be obtained by examining the amount of flux that results by reversible domain rotation. For this purpose, some hysteresis loops of the four materials described earlier are shown in Fig. 17. The scale chosen is such that the open portion of the hysteresis loop appears very small. The loops are taken by means of the familiar current-switching method, using the fluxmeter, or overdamped ballistic galvanometer. In all of the ballistic measurements to follow, a Leeds Northrop Type *HS No. 2285 f* reflecting galvanometer was used. This or its equivalent is recommended for its high sensitivity and long period. The fluxmeter made from this galvanometer has a reasonably slow return, making it a good integrator for flux measurements.[29] More refined experiments have recently been made using an integrator and some amplifiers in a Reeves analog computer with an Electronics Research Associates plotting table.

The difference in grain orientation of the chemically similar 48 Alloy and Orthonol is shown by comparing their respective curves in Fig. 17. In Orthonol

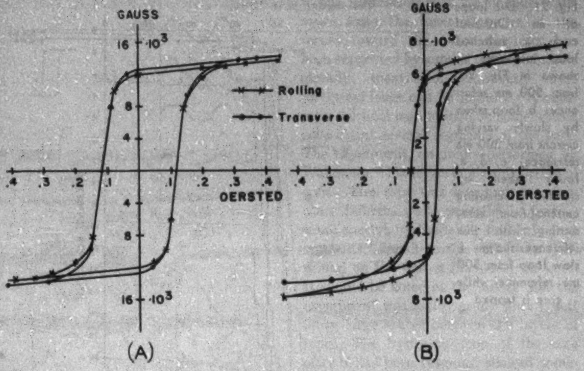


Fig. 18. D-c loops in rolling and transverse directions. A—Orthonol. B—Hy-Mu 80

it is seen that the remanent flux density is approximately 95% of the maximum, while in the 48 Alloy it is about 85%. The flux change in the "saturated region" is due to domain rotation. In the silicon-iron, the domain rotation is not complete until about 1,500 ampere turns per meter have been applied. In 48 Alloy about 800 ampere turns per meter are required. These large fields are a measure of the anisotropy energy required to rotate the domains. For HyMu-80, the rotation is complete with the application of about 150 ampere turns per meter. The comparison between the values of field required for rotation show that the anisotropy energy for the 80-20 nickel-iron alloy is much smaller than that for the other alloys, as expected from the figures in Table I.

Further confirmation of the degree and direction of orientation of the various tapes can be obtained by comparing toroidal cores made from tape cut parallel to the direction of rolling with cores made from tape cut perpendicular to the direction of rolling. The oriented 50-50 nickel-iron, for example, has its crystallites oriented with one cube edge parallel to the direction of rolling, and another perpendicular to the direction of rolling, but in the rolling plane, as shown in Fig. 14. Since the direction along the cube edge is the easy direction of magnetization, the material should magnetize with equal ease along the tape and across the tape. Thus, the two specially prepared cores should show substantially the same magnetic properties. Experiments have been performed by Wakeman[25] on cores of this kind, and confirm the results expected. The special cores were made of 5 wraps of 1-inch tape, 2 mils thick, wound on a 0.312-inch-diameter bobbin. Measurements of the finished cores indicate a stacking factor of only 0.2, which means that only 20% of the volume occupied by the core is actually tape cross section. Thus, there is room for large differences in the fabrication of the cores, and confirmation of theory can be expected only to within the uniformity achieved.

Measurements made on the two special Orthonol cores, shown in Fig. 18(A), confirm our expectations. The difference in height of the loop can easily be accounted for by a difference in effective core area of the two samples.

In HyMu-80 the crystals are randomly oriented. Thus it would be expected that the hysteresis loops measured in the rolling or transverse directions would be the same. The loops measured are shown in Fig. 18(B). Again, the difference in effective areas of the two cores may account for the different heights of the loops. The shapes of the two loops are somewhat different, but this may be accounted for by the different distribution of air gaps wound into the two cores, especially since the cores were made from only five wraps of tape. Note that the hysteresis loops of Figs. 18(A) and (B) show only that the measurements in the rolling and transverse directions are the same. Their similarity does not support the explanation of why they are the same.

In 3-97 silicon-iron, X-ray patterns show that the orientation is not complete. The orientation that exists, however, is expected to be as shown in Fig. 12. Measurements are expected to show two things. First, the percentage of flux reversal that takes place in low fields before domain rotation begins should be lower than that shown by Williams[30] for single crystals. Secondly, the orientation perpendicular to the rolling direction in the rolling plane should be in the *110* direction, and the field required to produce maximum flux density in the cross direction should compare with that required in the rolling direction in the same manner as shown by Williams. The curves obtained for the tape specimens are shown in Fig. 19 with the curves for single crystals. The expectations are therefore borne out by experiment.

Probably the most convincing evidence of magnetization by domain-wall motion is found when measuring d-c hysteresis loops with a reflecting galvanometer connected as a fluxmeter. The standard operating procedure with the fluxmeter, as with the underdamped ballistic galvanometer, is to establish a reference field, well into the saturated region, and then to switch over to a preset value of magnetic intensity on the other side of the loop. The galvanometer deflection then indicates the integral of the voltage induced in a coil on the core, or the flux increment from the reference point. Fig. 20 shows the experiment schematically. The outer loop in Fig. 21 shows the hysteresis loop obtained in this manner for a typical Orthonol core. The core is a Magnetics, Inc., *50039-1A* core made of 1-mil Orthonol, with an inner diameter of 1.25 inches and an outer diameter of 1.50 inches, and an effective cross-sectional area of 0.160 square cm (centimeters) according to published catalog data. It is observed during the course of the experiment that for points along the steep sides of the loop, the total galvanometer deflection proceeds rather sporadically and

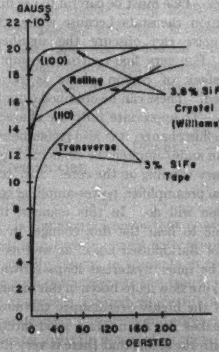


Fig. 19. D-c loops of 3% silicon-iron in rolling and transverse directions compared with loops from single crystals reported by Williams

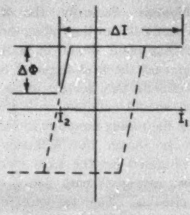


Fig. 20. Schematic representation of flux-current measurements

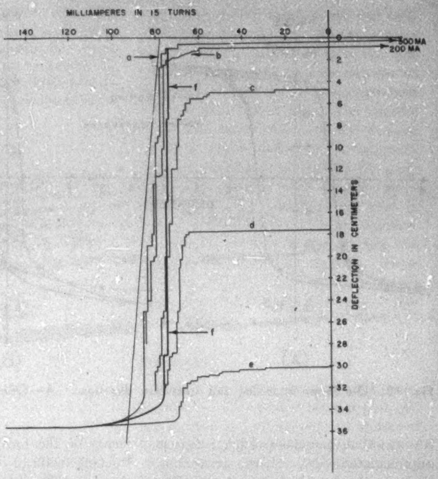


Fig. 21. D-c loops of an Orthonol core: a. switched loop measured as shown in Fig. 20, from 500 ma reference; b. loop taken by slowly varying current from 200 ma reference; c, d, e, loops taken by slowly increasing current from zero, starting with flux references shown; f. slow loop from 500 ma reference while core is tapped

may take a second or more to come to a stop. Then during the next few seconds, the deflection may increase in short jumps accumulating as much as 10% additional deflection. It is also found that when the experiment is repeated several times for a given value of exciting current, the indicated flux deviates from 10 to 25%, and occasionally as much as 50%, from the mean of a series of measurements. The experiment, in other words, is not repeatable. Thus, the steep portion of the d-c loop is not a single distinct curve, but is actually a diffuse region whose density indicates the probability of finding the value of flux for the given value of current.

It is very plain from this experiment, that the flux change takes place by means of domain-wall motion as described earlier. But the demonstration is even more convincing if one measures slow, continuous-current hysteresis loops. That is, instead of switching the exciting current to a preset value, as described in Fig. 20, the exciting current may be brought up from zero by slowly adjusting a sequence of rheostats in series with the battery. One must be careful to use high-resolution rheostats because sliding-contact noise can obscure the significant data. For very fine control, a compression stack of carbon plates is recommended. These can be obtained commercially. To appreciate fully the character of the flux change, it is also recommended that an audio amplifier be connected to an auxiliary winding on the core. Any commercial preamplifier, power-amplifier combination will do. In this manner, it is possible to hear the flux change, in the form of Barkhausen noise, as well as see it. The inner hysteresis loops shown in Fig. 21 are slow loops taken in this manner. Use of the highly overdamped galvanometer makes the fluxmeter a good integrator, with the result that there is very little return of the reflecting coil during the course of the slow measurements. The loops shown were recorded in 1 to 2 minutes. During the latter part of each loop, a drift of approximately 1 cm was accumulated, and has been corrected in the loops shown. Use of the electronic computer removes the drift problem.

For all of the loops shown in Fig. 21, the magnetic intensity is cycled between plus and minus 7 times the coercive force. The actual value used was 75 ampere turns per meter. These points become reference levels, from which all measurements are made. In each case, the core is left at the negative reference point, minus 75 AT/m, and the magnetic intensity is raised to varying positive values, saturating the core to different degrees. Following this, the magnetic intensity is reduced to zero and increased in the negative direction, thus measuring the starting field corresponding to the different levels of prior saturation. The negative halves of the loops are shown in Fig. 21. For part *a*, the saturating magnetic intensity is the positive reference value, or 7 times the coercive force. For *b*, the saturating magnetic intensity is 30 AT/m, or about 3 times the coercive force. For *c*, *d*, and *e*, the flux is not permitted to return all the way to positive saturation, but is arrested by means of the current controls. The values of saturating magnetic intensity for these three loops are only slightly different. Thus the starting fields are seen to be almost equal for these three conditions. For the conditions represented by *a* and *b*, however, the inner portion of the core is subjected to magnetic intensities of about 7 and 3 times the positive magnetic intensities represented by *c*, *d*, and *e*. The consequent increase in starting field is evident from the curves.

Once the magnetization has started to change in large steps, it takes very little to move the domain wall through the material. An interesting loop for the same core is shown in Fig. 21*f*. This loop differs from the loop of *a* in that the core was continually tapped with the eraser end of a pencil while the flux change was in progress. The effect can be described as follows: During the flux reversal, the domain walls proceed outward until they encounter some geometrical irregularity. The flux reversal continues if the current is increased, as shown in the other loops. But it is also possible to dislodge the domain walls from these irregularities, without increasing the current, by tapping the core. In many cases, this brings about further large spontaneous jumps in the flux. The loop shown in Fig. 21*f* is one in which tapping the core produced a loop with very straight sides. The large discontinuities shown are actually accumulated in several shorter steps. The discontinuities shown could have been made longer by tapping the core harder.

It might at first be puzzling to note that negative saturation is approached with different values of magnetic intensity in these several loops. This fact, however, gives us good confirmation that the magnetization takes place from the inner surface of the core outward. In each case, reading from *a* to *e*, saturation is approached by the motion of domain walls closer to the inner diameter of the core. It is expected, therefore, that the corresponding magnetic intensities would be decreased for this sequence of loops.

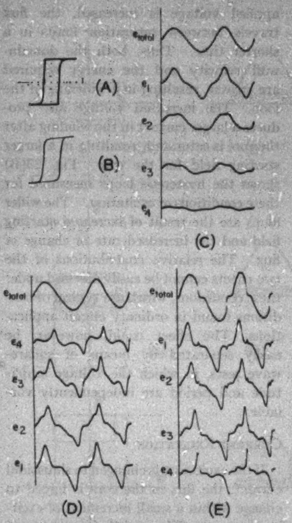


Fig. 22. 400-cycle hysteresis loop on Orthonol core: (A) Before and (B) after search coils were wound on core segments. (C, D, E) waveforms on induced voltage on cores segments for 50%, 100%, and overexcitation. Segments are numbered 1–4 from inside to outside

As explained in connection with Fig. 9, the magnetomotive force corresponding to saturation must be above the value required for domain nucleation because of the increase of domain-wall length in the outer portion of the core. In Fig. 21 a line of this slope is shown, and is seen to correspond approximately to the slope of the measured loop. The nature of the experimental flux reversal, however, indicates that the actual coercive force of a single crystal would be considerably below that of the tape-wound core. Thus the shape of the loop in detail is primarily determined by the difficulty with which the domain walls move through the tape structure. The most convincing evidence of this fact is the loop f of Fig. 21, in which the core was tapped. Since the core properties do not change drastically from the inner to the outer diameter, however, it is expected that the increase in magnetomotive force between the starting value and the saturation value would correspond closely to the increase predicted from the simple calculation indicated in Fig. 9.

The contention that the domain walls move from the inner diameter of the toroid toward the outer diameter has been supported by some of the experiments already mentioned. However, more direct evidence can be found by rewinding a toroidal core to include some search coils about several segments of the core. The experiment, originally performed by Wakeman,[25] made use of a 1-mil Orthonol core. The tape was rewound onto another bobbin, while periodically placing small spacers to divide the core into four segments. Small search coils were then wound on the four segments of the core. It should be noted at the outset that the experiment was rather crude since it left fairly large spaces between the layers of tape. The hysteresis loop of the core after it had been rewound showed somewhat less squareness. However, there was every reason to believe that the core would still magnetize from the inner diameter outward.

The core was excited with a 60-cycle sinusoidal voltage. The voltage induced in the four search coils were then recorded in a multiple-exposure photograph. Photographs were made for the core at full excitation, and about half excitation. The results of the experiment are shown in Fig. 22. In Fig. 22(A), hysteresis loop of the same size but unaltered core is shown. Fig. 22(B) shows the hysteresis loop of the core rewound with spacers and search coils in place. Fig. 22(D) shows the voltages induced in the four segments of the core compared to the total induced voltage with enough excitation to saturate the core fully. Fig. 22(C) shows the same voltages with about half excitation. In Fig. 22(C) and Fig. 22(D), it is obvious that the flux reversal proceeds from the inner segment of the core outward. In Fig. 22(D), where the core is only partially excited, the flux changes are fairly well limited to the inner segments of the core.

If the domain wall in the core were a simple cylinder, as imagined for the single crystal, the induced voltage in three of the search coils would always be zero, while in the fourth, the induced voltage would be equal to the total applied voltage. This induced voltage would switch from coil to coil with much more well-defined switching times than those shown. The photographs, therefore, indicate that the simple construct is not valid for the tape-wound core. But it must be recalled that some fairly large spaces have been left in the core, and these spaces probably contain some reverse domains when the core is saturated. Thus, we would expect the overlapping of induced voltages on the search coils, but it is not known whether the effect would be as large as that measured. A somewhat more refined experiment should provide at least a partial answer to this question.

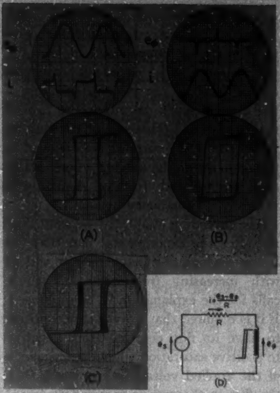


Fig. 23. Hysteresis loops of core under varying conditions of excitation

A—Sinusoidal voltage
B—Sinusoidal current
C—Increasing overexcitation with sinusoidal voltage
D—Circuit connections

An interesting result of this experiment is shown by the small peaks of induced voltage in the search coils at the end of each half-cycle, seen in Fig. 22(E). These are irreversible flux changes that occur only at the end of the half-cycle when the core saturates and the magnetomotive force becomes large. Note that they do not appear in Fig. 22(D) where the core does not saturate. If these voltage pulses were due to domain rotation, there would follow each positive pulse a negative pulse of approximately equal magnitude.

Hysteresis Loops at Power Frequencies

Width of the Loop

It is commonly known that power loss in an iron core increases with frequency. As usually taught to students in connection with power circuits and transformers, the theory is limited to justifying the dependence of hysteresis and eddy-current losses on a certain power of frequency, and graphical methods for separating out the components. Thus, most engineers are imbued with the notion that there is something inherent in the frequency that affects the power loss in a magnetic core. However, magnetization in these cores takes place by domain-wall motion and domain rotation, which are not homogeneous processes. The velocity and area of the domain walls determine the rate of magnetization. Throughout the flux reversal, the eddy currents are confined to the region near the moving domain walls, and are related to the domain-wall velocity, not frequency.

The preceding qualitative account does not alter the experimental fact that the hysteresis loop of a magnetic core widens with increasing frequency. It merely calls for the explanation to be rephrased. If it is required that a given core continue to have its flux changed from negative to positive saturation while the frequency of the applied voltage is increased, then the flux in the core is required to travel between the saturation limits in a shorter time. This requires that the applied voltage be increased proportionately. The domain-wall velocity must increase, resulting in higher eddy-currents and consequently higher power loss in the core.

Suppose that an experiment to show the widening of the loop were to be carried out by increasing the frequency by a factor of 2. The loop in the test core would widen as expected because the domain walls would have to travel twice as fast. The number of volts per turn on the winding would also double, thus requiring twice the applied voltage. If, for comparison, we take a core of the same tape width and the same inner diameter, but with only half the radial build-up, in order to have just half the cross-sectional area, the same applied voltage at the higher frequency would move the domain walls at the same velocity as before, and would still saturate the core at the end of the half-cycle because of the smaller cross-sectional area of the core. Since the velocity of the domain walls is unchanged, the loop has the same width as the larger test core at the lower frequency, but the loop is only half as high if total flux is measured. It is assumed that the ID/OD ratio is large in both cases. Thus, for a given wall velocity, the power loss per unit volume is the same, regardless of frequency. It is seen from this argument that the dependence of loop width on frequency of excitation is the result of core geometry and magnitude of applied voltage. Frequency itself is not directly significant.

VOLTAGE EXCITATION

The shape of the hysteresis loop usually depends upon the combination of several effects. These effects may be qualitatively explained in terms of the model that we have developed. The difference between the hysteresis loops obtained with sinusoidal voltage and sinusoidal current excitation is shown in Fig. 23. Sinusoidal voltage excitation is obtained by making the product of the circuit resistance and the exciting current very small compared to the voltage applied. The sinusoidal current condition is obtained by choosing a winding and series resistance so that the induced voltage in the reactor is small compared to the product of the exciting current and the circuit resistance.

In the case of sinusoidal voltage excitation, the rate of change of flux is prescribed and the current that flows is a measure of the velocity of the domain walls and thus the energy required to reverse the flux at the prescribed rate. The relatively slow establishment of domains of reverse magnetization in the core results in the rise of exciting current to values somewhat in excess of the static starting fields shown in Fig. 21. Subsequent rapid change of flux in the core generates a voltage e_φ that opposes the supply voltage e_s as shown in Fig. 23(D). The current that flows equals the difference between the two voltages divided by the circuit resistance. The result is a dynamic hysteresis loop that is somewhat wider and more rectangular than the d-c loop.

With a sinusoidal voltage applied, very nearly the full half-cycle is available for the flux to travel between its saturated extremes, as shown by Fig. 23(A). If the applied voltage is increased, the flux travels between saturation limits in a shorter time. Thus, both the domain-wall velocity and the energy required are greater, resulting in a widening of the loop. The increased voltage also produces a larger current in the winding after the core is saturated, resulting in a larger starting field for the core. Fig. 23(B) shows the hysteresis loops measured for these conditions of excitation. The wider loops are the result of increased starting field and the increased rate of change of flux. The relative contributions of the two effects cannot be easily assessed under these conditions, which are typical of conditions found in ordinary circuit applications. The effect could, however, be easily separated by means of square-wave tests, in which the voltage amplitude and period are independently variable.

CURRENT EXCITATION

If the excitation is changed to sinusoidal current, the flux in the core is forced to change within a small increment of exciting current. The small increment of exciting current corresponds to a short interval of time. As expected, therefore, the hysteresis loop is somewhat wider than it is for sinusoidal voltage excitation. This is illustrated experimentally in Fig. 23(C).

MINOR LOOPS

Probably the most interesting and to many the most puzzling property of rectangular-loop materials is the fact that minor loops do not "cluster" about the origin, the way they do in lower grade materials. Sets of minor loops for Orthonol, 48 Alloy, HyMu-80, and Magnesil are shown in Fig. 24. The loops cluster about positive saturation in all of these pictures except for Magnesil, the 3% to 97% silicon-iron alloy, which shows no visible sign of a symmetry. The loops may also cluster about negative saturation, of course, if the proper initial conditions are satisfied. Generally in an experiment of this kind, there is no way of knowing at which end the minor loops will be grouped. The excitation voltage used to obtain Fig. 24 is coupled through a transformer, and consequently there is no d-c component present.

The observed clustering of minor loops at one or the other saturation limits can be explained in terms of a set of curves such as those shown in Fig. 21. It is assumed that as higher degrees of saturation are reached, measuring from negative saturation, higher starting fields and therefore exciting currents are required.

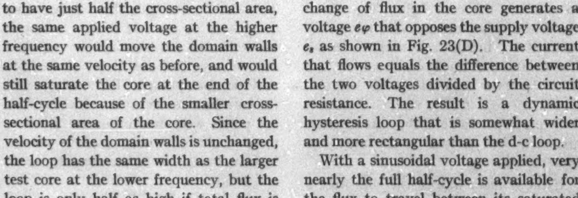


Fig. 24. Clusters of, minor loops of: A—Orthonol. B—HyMu-80. C—48 Alloy. D—Magnesil

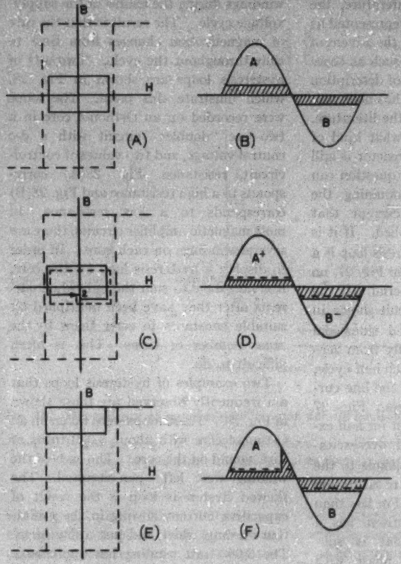


Fig. 25. Mechanism by which minor loops migrate toward saturation

The reasoning behind this assumption is still open to question. Until recently it was believed that the widening of the loop is due to the increased nucleating field with approach to saturation. Observed minor loops under dynamic conditions support this point of view, but it is now believed that the correct explanation is much more complicated. In any event, measurements show that the loop widens with the approach to saturation.

We shall assume that a sinusoidal voltage is applied to a reactor in series with a resistor, as shown in Fig. 23(D). It is assumed that the resistance is small enough so that most of the applied voltage appears across the reactor. The voltage equation for the circuit is

$$e_s = iR + e_\phi \tag{1}$$

It also follows that equation 1 can be integrated over 1 cycle of the supply voltage without destroying the equality.

$$\int_0^T e_s dt = \int_0^T iR dt + \int_0^T e_\phi dt \tag{2}$$

If it is assumed that a symmetrical minor loop exists and is centered about the origin, as shown in Fig. 25(A), the integral of each term in equation 2 is zero separately. The flux in the core and the exciting current are periodic with the period of the supply voltage. The third term in equation 2 represents the increment in flux linkages in the core during the period.

$$\Delta N\phi = \int_0^T e_\phi dt \tag{3}$$

In Fig. 25(B) the terms of equation 2 are shown graphically for one cycle of the supply voltage. The shaded area is the integral of the voltage dropped across the resistor. The steady-state requirement that the flux be periodic means that the unshaded areas *A* and *B* must be equal, since they are numerically proportional to the positive and negative flux increments.

If it is assumed, in contrast, that the hysteresis loop is displaced upward from the origin, with no change in applied voltage, the conditions of symmetry are no longer found. Because of the displacement, the negative exciting current is larger in magnitude than the positive exciting current. Thus, the resistive voltage drop for positive flux increments is smaller than it is for negative flux increments, increasing the area *A* and decreasing the area *B*. This results in a larger positive flux increment than negative flux increment on each succeeding half-cycle. A diagram of such a loop is shown in Fig. 25(C), with the corresponding voltage and current waveforms shown in Fig. 25(D). From equation 2, the net increment of flux linkages in each cycle is given by the relation

$$\Delta N\phi = \int_0^T e_\phi dt = -\int_0^T iR dt \tag{4}$$

The circuit reaches the steady state once more when the net increment of flux during a cycle is zero, or when the integral of the exciting current is zero. But as the transient proceeds, the current during the positive half-cycle decreases and the current during the negative half-cycle increases. Thus the equilibrium condition can be reached only when the core saturates, Fig. 25(E) and Fig. 25(F), thereby adding the necessary positive contribution to the current integral. The waveforms of the Fig. 25(F) show that the areas *A* and *B* are equal once more.

Two equilibrium states have been described, one centered about the origin, Fig. 25(A), and the other at positive saturation, Fig. 25(E). A third equilibrium state exists at negative saturation. When the loop is in the equilibrium state centered about the origin, a displacement in either direction causes the loop gradually to move toward positive or negative saturation. Thus, it is expected that the equilibrium state centered about

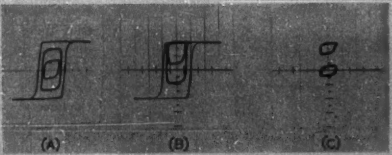


Fig. 26. Hysteresis loops for Orthonol core showing

A—Symmetrical minor loops

B—Asymmetrical minor loops

C—The change from a symmetrical to asymmetrical loop with time

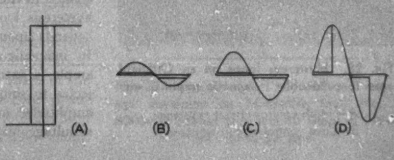


Fig. 27. Voltage-current waveforms for rectangular-loop saturable reactor showing variation of fundamental component of current with applied voltage

the origin is unstable, and the other two are stable. Ordinary experiments indicate that this is true. However, with some care it is possible to establish symmetrical loops and maintain them for some time. In the large majority of cases, the symmetrical loops persist for only fractions of a minute, and frequently for only a few seconds. On one occasion, however, a small symmetrical loop has been maintained for over an hour in an Orthonol core. Usually, switching transients on the power line are enough to upset the balance. It does appear, however, that there is a small range of stability for the symmetrical loops.

Fig. 26(A) shows a set of these symmetrical loops for an Orthonol core. Fig. 26(B) shows the usual asymmetrical group for the same core. Fig. 27(C) shows a small symmetrical loop in the core immediately after it was established, and a few seconds later after it had migrated toward positive saturation. The vertical displacement of the loop does not come through the electronic integrator, but must be put in to correspond to the saturation flux in Fig. 26(A). It is commonly found from experimenting with cores of this kind that it takes from 1 to 10 seconds, and sometimes longer, for the symmetrical loop to travel to one saturation limit or the other. This indicates that the flux increment per cycle is extremely small, which indicates a very small circuit resistance or a very slight widening of the loop with displacement, or both.

Circuit Aspects

The applications of tape-wound cores are many and varied and are rapidly increasing in number. It would not be possible to show in extensive detail those properties of circuits that relate directly to the magnetic phenomena discussed in the preceding paragraphs. There are, however, a few simple experiments that serve to illustrate some of the common circuit manifestations.

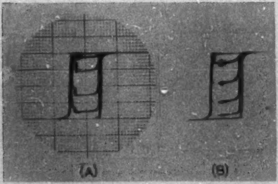


Fig. 28. Hysteresis loops on an Orthonol core in a "doubler" magnetic amplifier with

A—High control-circuit resistance
B—Low control-circuit resistance

In much of the early literature, the saturable core reactor was represented as a variable reactance. With the advent of rectangular-loop materials, such as those we are discussing, this type of description became less successful, and has now vanished almost entirely from the literature. However, the question of what kind of element the saturable-core reactor is still arises upon occasion. This question can be easily answered by examining the hysteresis loop and the current that flows when a voltage is applied. If it is first assumed that the hysteresis loop is a perfect rectangle, as shown in Fig. 27, no energy is stored in the material. If the reactor is placed in the circuit shown in Fig. 23(D) and the applied sinusoidal voltage is increased gradually from zero until the core saturates on each half-cycle, the amplitude and phase of the line current vary over wide ranges. Fig. 27 shows the voltage and current for half excitation, full excitation, and overexcitation. Full excitation corresponds to the maximum flux excursion before saturation occurs on both half-cycles. For less than full excitation, the line current has a fundamental component that is substantially constant, and in phase with the applied voltage. From the standpoint of energy, therefore, the reactor is purely dissipative and generates harmonics in the line current. As the voltage increases, the apparent resistance increases since the fundamental component of line current remains unchanged. For overexcitation, shown in Fig. 27(D), a large component of line current appears that is nearly 90 degrees lagging the line voltage. Thus we have a nonlinear circuit component that draws a lagging current from the line but stores no energy. As the voltage is increased further, the angle of the fundamental component of current again approaches zero. It should be obvious from this experiment that representation of the saturable-core reactor as a variable linear reactance is not possible. From our earlier discussion, we recognize that there is some energy stored in the material in the form of domain walls and domain rotation. Thus the hysteresis loop of Fig. 27(C) is not accurate. But the general circuit behavior of the reactor is not affected by this correction.

It should be apparent from our discussion in the foregoing sections that the hysteresis loop traced for a given core depends upon the manner of excitation. In magnetic-amplifier circuits, for example, a saturable-core reactor is subjected to sinusoidal voltages of different amplitudes and d-c voltages, switched simultaneously or singly onto the various windings during the course of the supply-voltage cycle. The result is that the rate of magnetization changes from time to time throughout the cycle. Two sets of hysteresis loops are shown in Fig. 28, which illustrate this point. The loops were recorded on an Orthonol core in a two-core "doubler" circuit with a d-c control voltage, and two values of control-circuit resistance. Fig. 28(A) corresponds to a high resistance and Fig. 28(B) corresponds to a low resistance. In most magnetic-amplifier circuits, there are several windings on each core. In order to display a hysteresis loop for the core, it is necessary to sum the individual currents after they have been multiplied by suitable constants to refer them to the same number of turns. This is often difficult to do.

Two examples of hysteresis loops that are frequently observed are those shown in Fig. 29. These loops were taken on an Orthonol core with about 3,000 turns of wire wound on the core. The ends of the winding were left open-circuited. The skewed hysteresis loop is the result of capacitive current flowing in the excitation winding, which had just a few turns. The 3,000-turn winding has appreciable interwinding capacity and acts as a capacitively loaded secondary on the core. To confirm this argument, the loop was duplicated by removing the 3,000-turn

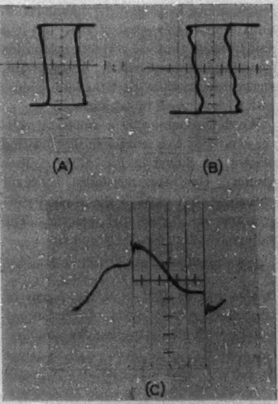


Fig. 29. Effect of winding capacity on 400-cycle hysteresis loop

A—With filtered power line
B—Same loop with line filter removed, showing 11th harmonic
C—Oscillation on load-current waveform caused by winding capacity

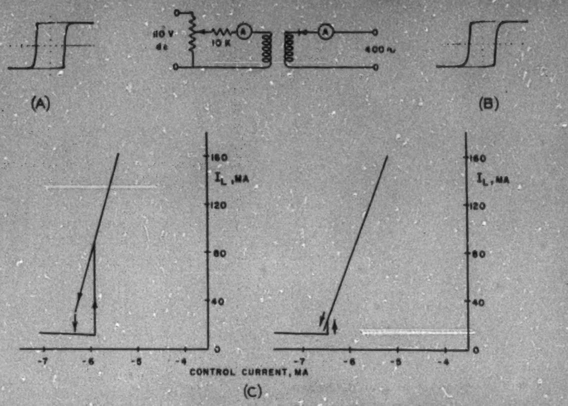


Fig. 30. Dependence of amplifier stability upon domain nucleation properties of core material

A—Core before stressing. B—Core after stressing. C—Control characteristic before stressing, showing large unstable region, and after stressing, small unstable region

winding and placing about 100 turns on the core in its place, with an ordinary capacitor connected to its terminals. The wiggles on the sides of the loop are the result of a 1% eleventh harmonic distortion in the 400-cycle generator. The harmonic has been exaggerated by the winding capacity. When the 400-cycle line was filtered, the loop of Fig. 29(A) resulted. Fig. 29(C) shows the effect of a large number of turns on the output current of a simple magnetic amplifier. A self-oscillation, or "ringing," in the load current is present because of the large winding capacity.

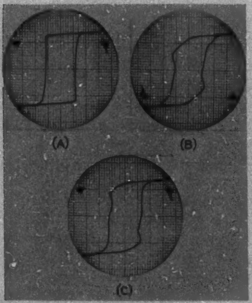


Fig. 31. Hysteresis loops on an uncased Orthonol core

A—Normal
B—Stress applied by squeezing between the fingers
C—With loose wraps introduced about one third of the way radially into the core, measuring from the inner diameter

The phenomena shown in Fig. 29 are not due to the magnetic properties of the core. They are mentioned here because capacitive effects may otherwise be wrongly attributed to the magnetic properties of the core.

Probably the most important effect of "second order" core properties on observed circuit behavior is the instability found in many of the high-gain magnetic amplifiers. As an example, Fig. 30 shows the instability obtained in a half-wave amplifier and the shape of the hysteresis loop for the corresponding core. The core material is Orthonol. The instability is due to the nature of the minor loops discussed in the previous section. In Fig. 30(A), the loop is seen to be very nearly rectangular, and the exciting current for the minor loops is markedly dependent upon the prior approach to saturation. The analysis of the circuit, which includes this property of the dynamic hysteresis loop, predicts a wide range of instability in the static control characteristic. The measured characteristic is shown in Fig. 30(C) and confirms the prediction. When the hysteresis loop is less rectangular, the degree of instability is reduced, according to the analysis. The measured hysteresis loop and corresponding control characteristic are shown in Figs. 30(B) and (C). The experimental conditions are identical in all respects except for the shape of the hysteresis loop. In fact, the cores themselves are identical. The hysteresis loop is altered by placing a small screw clamp across the thickness of the core. Tightening the clamp squeezes the layers of tape together so as to change the hysteresis loop from that shown in (A) to that shown in (B).

The effects of stressing a core are most dramatically illustrated by Orthonol, which is the most stress-sensitive material of those discussed. An annealed core of 4-mil Orthonol without protective case has the hysteresis loop shown in Fig. 31(A) at 60 cycles. This is the expected loop. By squeezing the core into an oval shape, the loop of Fig. 31(B) results. Releasing the stress returns the loop to its former shape. The stress involved in this experiment is much larger than that applied in Fig. 30(B).

If the inner and outer portions of the core are then rotated with respect to each other to increase the tape spacing at a point within the core, the loop in Fig. 32 results. It is difficult to do this without stressing the core as well. But the kink in the side of the loop definitely shows the point at which the exciting current increases in order to magnetize the core beyond the diameter at which the loosened turns occur. With a little care it is possible to return the core to a condition that again yields approximately the hysteresis loop of Fig. 31(A).

Many recent contributions to the literature discuss the properties of magnetic cores in circuits. These contributions discuss the core properties from many different points of view, and give convincing evidence that a knowledge of magnetic materials is important to engineers responsible for circuit design.[32–36]

High-Speed Magnetization

High-speed flux reversal in magnetic materials is a subject of great current interest and research.[4–6,37–40] Nonmetallic ferrites and thin alloy films seem to be of greatest interest at the present time. Our discussion, however, is limited to tape-wound cores.

At high speeds, we still speak of mag-

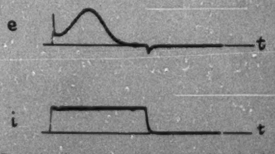
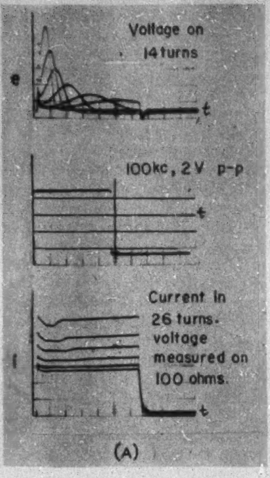


Fig. 32. Typical current and induced voltage waveforms on Magnetics, Inc., No./80004-B bobbin core with 20 wraps of 1/8-mil HyMu-80 tape



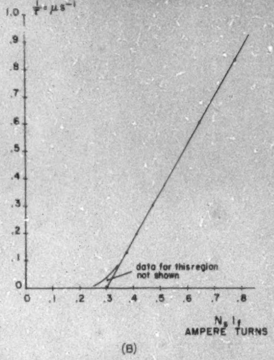


Fig. 33. A—Induced voltages on bobbin core of Fig. 32 with variation of current-pulse amplitude; B—Curve of inverse switching time versus magnetomotive force taken from photographs in A

netization by domain-wall motion and domain rotation. However, the dynamics of domain walls, ignored heretofore, must be explained in terms of the propagation of spin waves, which are expressed in terms of complicated quantum mechanical language. The viscosity of domain walls implies that large fields are required to reverse the flux in a short interval of time. Also, the bobbin cores used for high-speed switching have a small radial build-up and low space factor. Thus, it is probably correct to think of domains being created on all the surfaces of the tape throughout the core rather than at the inside diameter alone. These new factors may be used to modify the model of low-speed flux reversal to obtain a fairly satisfactory explanation of the observed magnetic properties under high-speed conditions.

REVERSALS BY DOMAIN-WALL MOTION

When a tape-wound bobbin core is excited with a rectangular current pulse the induced voltage on a test winding has the form shown in Fig. 32. The tape is 1/8-mil-thick HyMu-80. Two features of the voltage pulse must be explained. The first is the initial sharp peak, and the second is the longer duration pulse with a maximum somewhere near the middle. The current opinion seems to be the following.[37–40] The sharp peak in the voltage waveform is the result of fast but elastic domain-wall motion, which is completely reversible. The remainder is the result of domain walls expanding about their nucleating centers and combining with one another, with the maximum rate of domain-wall annihilation occurring at the voltage peak near the middle of the pulse. Increasing the current pulse has the effect of accelerating the entire process.

The initial sharp peak in the voltage waveform is supposed to be the result of reversible domain-wall motion against elastic forces. Experimentally, this is verified by applying a pulse of short duration corresponding to this portion of the voltage wave. Presumably, this does not allow sufficient time for the irreversible growth of domain walls. When the short current pulse is removed after about 0.1 microsecond, a voltage pulse of the opposite polarity is induced and area equal in magnitude to the positive pulse. Thus the reversibility of the process is established. But from calculations of domain-wall energy, Newhouse[38] argues that the range of reversible domain-wall motion is much shorter than necessary to account for the large reversible voltage pulse in question. In fact, measurements have shown that this reversible flux change can be made as large as from 20 to 50% of the total irreversible flux change. Thus other contributory mechanisms are indicated. The problem has not been solved as yet, but one obvious reversible contributor could be domain rotation. Whether or not this contribution would be sufficient to account for the additional flux change that takes place in this short period is questionable.

The explanation of the second maximum in the voltage pulse is that the regions of reverse magnetization expand about their nuclei until they begin to overlap. Overlapping of domains results in domain-wall annihilation and subsequent rapid changes in flux. This model of domain growth is the same as that used to explain the behavior of slow hysteresis loops, shown in Fig. 21. The time required for the domain growth differs, of course, but the qualitative similarity is apparent. One major difference in the mechanism is related to the high applied field and the low stacking factor of the bobbin core. The large fields applied for high-speed switching are from 10 to 50 times the coercive force of the material. Thus, the field is more than adequate to reverse the flux anywhere in the core. The concept of the domain walls sweeping outward radially after nucleating at the inner surface is no longer tenable. It is more likely that reverse domains grow about nucleating centers over the entire tape surface, spreading over the surface and penetrating into the tape.[39]

The time required for switching the magnetization in a core from one saturation limit to the other depends upon the field applied. Fig. 33(A) shows a series of rectangular current pulses and the corresponding induced voltages. As the current pulses increase in amplitude, the time required for switching reduces. The pulse amplitude in magnetic intensity is customarily plotted against the corresponding inverse switching time. Such a curve, taken from the data in Fig. 33(A), is shown in Fig. 33(B). The slope of the curve is often taken as the figure of merit for the particular core. It would be expected from our understanding of domain-wall motion that when current pulses just large enough to switch the core are applied, the switching time would vary. It is not surprising to find, then, that the lower part of the curve in Fig. 33 (B) is not repeatable, but defines a diffuse region.

It is not possible to give a complete account of high-speed switching in magnetic materials without including alloy films and ferrites. The present discussion however, is limited to tape-wound cores. Further information on films and ferrites may be found in the list of references.

References

1. APPLICATIONS OF NONLINEAR MAGNETICS, Herbert F. Storm. *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 77, July 1958, pp. 380–88.

2. SYMPOSIUM OF PAPERS ON FERROMAGNETIC MATERIALS. *Proceedings*, Institution of Electrical Engineers, London, England, vol. 97, pt. II, 1950, pp. 119–272.

3. SOFT MAGNETIC MATERIALS FOR TELECOMMUNICATIONS (book), **C. E. Richards, A. C. Lynch**, editors. Interscience Publishers, Inc., New York, N. Y., 1953.

4. CONFERENCE ON MAGNETISM AND MAGNETIC MATERIALS. *AIEE Special Publication T-78*, Oct. 1955.

5. CONFERENCE ON MAGNETISM AND MAGNETIC MATERIALS. *AIEE Special Publication T-91*, Feb. 1957.

6. CONFERENCE ON MAGNETISM AND MAGNETIC MATERIALS. *Journal of Applied Physics*, New York, N. Y., vol. 29, no. 3, Mar. 1958.

7. FERROMAGNETISM (book), **R. M. Bozorth**. D. Van Nostrand Company, Inc., Princeton, N. J., 1951.

8. MODERN MAGNETISM (book), **L. F. Bates**. Cambridge University Press, Cambridge, England, third edition, 1951.

9. FERROMAGNETISMUS (book), **R. Becker, W. Döring**. Julius Springer, Berlin, Germany, 1939; also Edwards Brothers, Inc., Ann Arbor, Mich., 1943.

10. THEORY OF ELECTRIC AND MAGNETIC SUSCEPTIBILITIES (book), **J. H. Van Vleck**. Clarendon Press, Oxford, England, 1932.

11. FERROMAGNETIC DOMAIN THEORY (book), **C. Kittel, J. K. Galt**. "Solid State Physics," Academic Press, Inc., New York, N. Y., vol. 3, 1956, pp. 439–534.

12. FERROMAGNETISM IN RELATION TO ENGINEERING MAGNETIC MATERIALS, **F. Brailsford**. *Proceedings*, Institution of Electrical Engineers, vol. 103, pt. A, 1956, pp. 39–51.

13. SPACE GROUPS AND THEIR REPRESENTATIONS (book), **G. F. Koster**. "Solid State Physics," Academic Press, Inc., vol. 5, 1957.

14. QUANTUM THEORY OF MATTER (book), **J. C. Slater**. McGraw-Hill Book Company, Inc., New York, N. Y., 1951.

15. THE QUANTUM PHYSICS OF SOLIDS, **W. Shockley**. *Bell System Technical Journal*, New York, N. Y., vol. XVIII, 1939, pp. 645–723.

16. D-SHELL VS. D-BAND MODELS, **C. Zener, R. R. Heikes**. See reference 5, pp. 216–24.

17. FERROMAGNETISM AND THE BAND THEORY, **J. C. Slater**. *Reviews of Modern Physics*, New York, N. Y., vol. 25, no. 1, Jan. 1953.

18. A SURVEY OF THE THEORY OF FERROMAGNETISM, **J. H. VanVleck**. *Ibid.*, vol. 17, 1945, pp. 27–47.

19. L'HYPOTHESE DU CHAMP MOLECULAIRE ET LA PROPRIETE FERROMAGNETIQUE, **P. Weiss**. *Journal de Physique*, Paris, France, vol. 6, 1907, pp. 661–90.

20. MAGNETIC DOMAIN PATTERNS ON SINGLE CRYSTALS OF SILICON-IRON, **H. J. Williams, R. M. Bozorth, W. Shockley**. *Physical Review*, New York, N. Y., vol. 75, 1949, pp. 155–78.

21. STRUCTURE OF METALS (book), **C. S. Barrett**. McGraw-Hill Book Company, Inc., second edition, 1952.

22. DOMAIN STRUCTURE OF PERMINVAR HAVING A RECTANGULAR HYSTERESIS LOOP, **H. J. Williams, Matilda Goertz**. *Journal of Applied Physics*, vol. 23, no. 3, Mar. 1952, pp. 316–23.

23. PROPERTIES AND POTENTIALITIES OF COLD-REDUCED NICKEL-IRON, **W. F. Randall, H. H. Scholefield**. See reference 3, pp. 225–32.

24. SOME METALLURGICAL AND STRUCTURAL FACTORS AFFECTING PROPERTIES OF SOFT MAGNETIC MATERIALS, **W. F. Randall, H. H. Scholefield**. *Proceedings*, Institution of Electrical Engineers, vol. 97, pt. II, 1950, pp. 133–40.

25. SUPERPOSED MAGNETIZATION IN MATERIALS WITH RECTANGULAR HYSTERESIS LOOPS, **C. B. Wakeman**. *Doctoral Dissertation*, Yale University, New Haven, Conn., May 1955; also *Technical Report 56-199*, Wright Air Development Center, Dayton, Ohio, pt. II, also available, *No. PB 131209*, Department of Commerce, Washington, D. C.

26. ASSESSMENT OF INHOMOGENEITY IN STRIPS OF HIGH-PERMEABILITY ALLOYS, **A. C. Lynch**. See reference 3, pp. 183–90.

27. NEW DEVELOPMENT IN ELECTRICAL STRIP STEELS, **N. P. Goss**. *Transactions*, American Society for Metals, Cleveland, Ohio, vol. 23, 1935, p. 511.

28. THEORY OF DOMAIN CREATION AND COERCIVE FORCE IN POLYCRYSTALLINE FERROMAGNETICS, **J. B. Goodenough**. *Physical Review*, vol. 95, no. 4, Aug. 1954, pp. 917–32.

29. ELECTRICAL MEASUREMENTS (book), **F. K. Harris**. John Wiley & Sons, Inc., New York, N. Y., 1952, chap. 8.

30. MAGNETIC PROPERTIES OF SINGLE CRYSTALS OF SILICON-IRON, **H. J. Williams**. *Physical Review*, vol. 52, 1937, p. 747.

31. FLUX REVERSAL IN MAGNETIC AMPLIFIER CORES, **F. J. Friedlander**. *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 75, July 1956, pp. 268–78.

32. THE INFLUENCE OF MAGNETIC AMPLIFIER CIRCUITRY UPON THE OPERATING HYSTERESIS LOOPS, **Harold W. Lord**. *Ibid.*, vol. 72, 1953 (Jan. 1954 section), pp. 721–28.

33. FLUX RESETTING CHARACTERISTICS OF SEVERAL MAGNETIC MATERIALS, **Hoobert Huhta**. *Ibid.*, vol. 73, May 1954, pp. 111–14.

34. CORE-RESET FUNCTIONS IN MAGNETIC-AMPLIFIER ANALYSIS. PART I—A CORE-RESET FUNCTION, **G. C. Feth**. *Ibid.*, vol. 77, Sept. 1958, pp. 503–09.

35. THE OPERATION OF MAGNETIC AMPLIFIERS WITH VARIOUS TYPES OF LOAD. PART I—LOAD CURRENTS FOR GIVEN ANGLE OF FIRING, **L. A. Finzi, R. R. Jackson**. *Ibid.*, vol. 73, July 1954, pp. 270–78; PART II—CONTROLLING THE ANGLE OF FIRING. THE TRANSFER CHARACTERISTICS OF AMPLIFIERS WITH LOW-CONTROL IMPEDANCE, **L. A. Finzi, R. R. Jackson**. *Ibid.*, pp. 279–88.

36. DYNAMIC CORE BEHAVIOR AND MAGNETIC AMPLIFIER PERFORMANCE, **L. A. Finzi, D. L. Critchlow**. *Ibid.*, vol. 76, May 1957, pp. 229–40.

37. MAGNETIC MATERIALS FOR DIGITAL COMPUTER COMPONENTS, **N. Menyuk, J. B. Goodenough**. *Journal of Applied Physics*, vol. 26, pt. I, 1955, p. 8; pt. II, p. 692.

38. THE UTILIZATION OF DOMAIN-WALL VISCOSITY IN DATA-HANDLING DEVICES, **V. L. Newhouse**. *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 45, no. 11, Nov. 1957, pp. 1484–91.

39. INFLUENCE OF PULSED MAGNETID FIELDS ON THE REVERSAL OF MAGNETIZATION IN SQUARE-LOOP METALLIC TAPES, **D. S. Rodbell, C. P. Bean**. See reference 4, pp. 219–27.

40. MODEL FOR NONLINEAR FLUX REVERSALS OF SQUARE-LOOP POLYCRYSTALLINE MAGNETIC CORES, **M. K. Haynes**. *Journal of Applied Physics*, vol. 29, no. 3, Mar. 1958, p. 472.

◆

Discussion

W. C. Bullock (Westinghouse Electric Corporation, Baltimore, Md.): Dr. Barker is to be commended for an excellent discussion of factors affecting the performance of tape-wound magnetic cores. Many years' experience with high-performance tape-wound cores, both in magnetic-amplifier design and manufacturing and in tape-wound core application engineering, have convinced me that the characteristics of cores so clearly presented by Dr. Barker, if more widely understood by device designers, would resolve many of the engineering problems associated with the use of high-performance tape-wound cores.

In the section entitled "Circuit Aspects," Dr. Barker discusses very briefly an instability frequently encountered in high-gain magnetic amplifiers which results from "core triggering." The reasons for core triggering are clearly explained as an instability of minor loops when located at other than saturation following phenomenological treatments of Batdorf and Johnson[1] and Finzi and Critchlow (see reference 36 of the paper). Unfortunately, Dr. Barker leaves one with the impression that the cure for core triggering is the choice of a less rectangular loop material or stressing the core to reduce squareness.

The choice of a less rectangular loop, for example, replacing Orthonol with 48 Alloy, will not completely eliminate core triggering and anyone who intentionally permanently stresses a high-performance tape-wound core may experience many other difficulties because the stressing of the core cannot be controlled.

The solution offered by Lynn, et al.,[2] is to increase the gating volt-seconds applied to the core so that the core can never operate on a minor loop. When push-pull circuits are used, it is frequently possible to balance the amplifier so that core triggering occurs in a region of the control characteristic which is less critical for the application intended.

A second point which should be clarified is the stressing of a core by first squeezing it into an oval shape and then releasing. Frequently, during the core anneal, if the insulating coating on the tape is not perfect, turns of the core may weld and the cooling of the core may result in stress. Squeezing the core, provided the elastic limit of the metal is not exceeded, will frequently break these welds and result in a more nearly square flux-current loop and lower losses.

REFERENCES

1. AN INSTABILITY OF SELF-SATURATING MAGNETIC AMPLIFIERS USING RECTANGULAR LOOP CORE MATERIALS, **S. B. Batdorf, W. N. Johnson**. *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 72, July 1953, pp. 223–28.

2. SELF-SATURATING MAGNETIC AMPLIFIERS (book), **G. E. Lynn, T. J. Pula, J. F. Ringleman, F. G. Timmel**. McGraw-Hill Book Company, Inc., New York, N. Y., 1960, pp. 184–87.

Harold W. Lord (General Electric Research Laboratory, Schenectady, N. Y.): Although Dr. Barker does not definitely describe just how the domain walls travel from the inner diameter to the outer diameter of a toroid, one gets the impression that it may be by means of a domain boundary moving substantially lengthwise of the tape. This discussion is submitted to show that such a concept is not a valid one.

It is interesting to compare the average velocity of a domain wall that would be required by such an interpretation of the propagation within a tape-wound core with direct experiments on unwound tapes.[1] While the dimensions of this particular core are not given, if they are similar to those of other cores Dr. Barker employed, the total tape length is approximately 10 meters. Since the experiment was made at 400 cycles, the average velocity inferred from this interpretation is about 8×10^3 meters/sec. The velocity (V) of domain wall propagation is found to be of the form: $V = A(H - H_o)$, where H is the applied field, H_o is a sticking field that is very nearly equal to the d-c coercive force, and A is a characteristic mobility constant. Since, for 0.001-inch Orthonol, the excess of H over

H_o at $B = 0$ is less than 0.2 for a 400-cycle sine-flux excitation, and since the maximum dB/dt is average $dB/dt \times \pi/2$, the mobility given by this interpretation is greater than 60×10^3 meters/sec-oersted.

The fact that the maximum velocity inferred by a propagation interpretation of the experiment cited by Dr. Barker is 30 times that observed on an unambiguous experiment seems to indicate that one should view with some reserve any conclusion that propagation takes place lengthwise along the tape in a wound core.

It is also interesting to consider how such an interpretation would affect the measured properties of toroidal cores. Suppose we have one core which has twice the number of wraps as another core of identical core material, mean turn diameter, strip width and thickness. Then, if the magnetization proceeds by propagation along the strip length, the velocity in the larger area of core for a given excitation frequency will be twice that for the smaller core. With the same mobility for both cores, the value of $H - H_0$ observed for the larger core will be double that for the smaller core. Again, suppose the two cores are identical in all respects except that one has twice the mean diameter of the other one. Again, the value of $H - H_0$ observed for the larger core would be double that for the smaller one if magnetization proceeded in this manner.

Since observed changes in magnetization characteristics of tape-wound cores do not follow the pattern described above, it seems clear that the reversal of magnetization in tape-wound cores does not take place by the motion of a domain wall along the length of the tape.

Alternative explanations of the experimental results cited by Dr. Barker include the simplest one that for a given current the field is greater on the inside than on the outside of the core. Secondly, the process of rewinding the core on another bobbin, with spacers between four segments of the core to accept pickup coils, introduced strains as indicated by the change in hysteresis loop shapes shown by Figs. 22(A) and (B) of the paper. If such strains were greater on the outer turns than on the inner turns, the coercive force of the outside turns would be higher than that of the inner turns. Such a strain distribution could produce the observed results shown by other portions of Fig. 22.

I wish to acknowledge the assistance of Dr. C. P. Bean, of the General Electric Research Laboratory, in the preparation of this discussion.

REFERENCE

1. MOTION OF FERROMAGNETIC DOMAIN WALLS IN 48 NI-FE TAPES, Richard W. Cole. *Journal of Applied Physics*, New York, N. Y., vol. 27, no. 9, Sept. 1956, p. 1104.

R. C. Barker: The discussions of Mr. Bullock and Mr. Lord are most welcome. Mr. Bullock's first comment discusses the phenomenon of amplifier triggering as it is related to core triggering, in which a minor loop of a certain height may migrate from one end of the major loop to the other. My comments on this well-known experimental fact were directed primarily toward the core behavior. The little experiment that I presented makes use of a very rudimentary circuit so as not to confuse circuit problems with core problems. The use of stress applied to the core was intended as evidence that this is indeed a core problem, and was in no way meant as a suggested remedy. I will confess, however, that the application of stress has been discussed as a potentially useful technique, but has not been developed further to my knowledge.

As Mr. Bullock points out, amplifiers having a tendency to trigger as a result of migrating hysteresis loops can often be stabilized by making the applied voltage large enough to insure that each core reaches positive saturation on its gating half-cycle. It also happens at times that certain amplifiers using these cores show some triggering while the cores themselves do not. In such cases, the causes of triggering are a bit elusive, but often they are traceable to other core properties in conjunction with particular circuit connections. Little literature on these problems is available because engineers have a natural aversion toward lengthy discussions of circuits that do not work. I am happy to read Mr. Bullock's second comment on the stressing of cores to break welds, a phenomenon which I failed to include in my discussion.

In this paper, I have purposely avoided detailed descriptions of the magnetization process, which accounts for the inference that Mr. Lord has drawn regarding the structure of domain walls in tape cores. Since the same question has been raised by others, the best reply is probably a slight enlargement on the concepts behind the notion that the flux reversal sweeps from the inner diameter of the core outward. The fundamental reason behind the assertion is that, because of the geometry of core, the inner portion is always subjected to a field higher than the outer portion, as argued in the text discussing Fig. 9. Secondly there is the unmistakable experimental fact that even at 400 cycles the flux in the inner portion of the core reverses on a time schedule that is clearly ahead of that for the outer portion of the core. Several sophisticated experiments have been designed to confirm this fact.

The experiment associated with the data of Fig. 22 was a very crude one performed several years ago and reported here largely because it is a fairly obvious experiment to perform on a single core, and can be repeated by almost anyone who wishes to invest an afternoon's time.

The fact that the schedule of flux reversal of the inner portion of the core is ahead of the schedule for the outer portion means that at any time in the proceedings one can describe the inner portion of the core as having substantially reversed and the outer portion as not yet significantly reversed. Between these two regions is what is called a *transition region*, roughly cylindrical in cross section and with a diameter that increases with time. The purpose of recent research at this laboratory has been to perform experiments relevant to this concept in order to develop the details of the structure of the transition region, its mode of motion, and their dependence upon the rate of flux reversal. The fairly obvious experimental facts that Mr. Lord cites contribute to the description of the transition region; they do not refute its existence. The notion that the transition region is a simple wall across the tape and moving lengthwise along the tape is, of course, naïve. Even the notion that the transition region is a complex domain-wall structure essentially fixed in shape but moving at a velocity dependent upon the applied filed is untenable. It would predict quite different results for experiments similar to that of Fig. 22, and would also lead to the kind of results, cited by Mr. Lord, that are not experimentally observed. It would also result in a widening of the 400-cycle loop that would be seven times as large as that of the 60-cycle loop, a fact that is found to be incorrect by reference to any commercial core catalog.

These experimental facts have been evaluated in terms of the Sixtus-Tonks equation, cited by Mr. Lord, which has been shown to be valid experimentally in the stretched wires of Sixtus and Tonks,[1] Dijkstra and Snoek,[2] and the stretched tapes of Cole discussed in the very brief letter referenced by Mr. Lord. But the quantitative and even the qualitative results of the Sixtus-Tonks experiment cannot be obtained unless the specimen is under tension.[2] As Mr. Lord points out in his last paragraph, tension in this type of material increases its coercive force. When the tension is removed, it is expected from the experiments on wires that the entire tape will reverse more or less uniformly, and that no such linear velocity will be measured. But from either electrical or magnetic considerations, the speed of the flux reversal surely depends upon the applied field. The functional relationship between this speed and the applied field is indicative of the actual domain wall area active in the reversal.[3]

The results of many experiments of the sort recently published[4] at slow flux reversals, and other to be published soon have developed the concept of the transition region as consisting of domain nucleation and growth about nucleation sites, domain-wall annihilation, and co-ordinated domain-wall growth taking part in proportions that depend upon the speed of reversal and the applied field. It is clear experimentally that at increasing reversal speed the field increases, but not in proportion to the speed. But long continuous walls that sweep out a given volume of material in a unit of time do so with much less active wall area than wall growth about a number of nucleating sites in the same volume. Thus, it is inferred that at the higher speeds the reversal is less the result of long walls moving with higher velocity, and increasingly the result of domain growth about a fairly dense distribution of nucleating sites, probably at the tape surfaces. The increased field leads naturally to the conclusion that the transition region spreads out through the core as the speed of reversal increases. A transition-region velocity along the length of the tape can be defined which is a sort of phase velocity resulting from the growth and subsequent annihilation of walls about nucleating sites, giving the appearance of a highly complicated domain-wall structure of continually changing shape, moving outward through the core, or lengthwise along the tape. The degree of domain-wall correlation brought about by magnetostatic fields between wraps is not known, and most of the other microscopic details of this picture have not been directly confirmed experimentally. But all of the microscopic experiments performed to date support this view.

I wish to acknowledge the excellent experimental work performed during the past two years by Richard M. Brownell in the pursuit of a Doctorate degree at Yale University. These experiments have made signal contributions to the state of knowledge in this field.

REFERENCES

1. PROPAGATION OF LARGE BARKHAUSEN DISCONTINUITIES, K. J. Sixtus, L. Tonks. *Physical Review*, New York, N. Y., vol. 43, 1933, pp. 931–40.
2. PROPAGATION OF LARGE BARKHAUSEN DISCONTINUITIES IN FE-NI ALLOYS, L. J. Dijkstra, J. L. Snoek. *Philips Technical Review*, Eindhoven Netherlands, vol. 4, 1949, pp. 334–56.
3. DOMAIN-BOUNDARY CONFIGURATIONS DURING MAGNETIZATION REVERSALS, J. J. Becker. *Journal of Applied Physics*, New York, N. Y., vol. 30, 1959, pp. 387–90.
4. DIRECT-CURRENT MEASUREMENTS ON TAPE-WOUND CORES, R. C. Barker, R. M. Brownell. *Ibid.*, vol. 31, 1960, pp. 243S–44S.

Pigmented Polyester Magnet Wire for High-Temperature Operation

F. A. SATTLER
NONMEMBER AIEE

L. C. SCALA
NONMEMBER AIEE

Synopsis: This work was done to develop a conductor insulation with good space factor and winding properties, capable of operating at 1,000 degrees Fahrenheit (538 degrees centigrade). A flexible, clean-burning polyester resin was used as a binder for a glass frit and a refractory in a wire enamel composition. The polyester resin contributes sufficient flexibility to the enamel to wind it into equipment. After completion of the winding operation, the organic portion of the magnet wires is burned out and the glass frit is fused to bind the refractory. The resulting insulation is not a continuous film, and air breakdown values are obtained.

POLYESTER resins based on isophthalic and terephthalic acids were investigated as binders since they burn clean without leaving a carbon residue and can be formulated to high flexibility for maximum pigment loading. A low functionality phthalic acid isomer polyester resin gave the best physical properties.

This resin, dissolved in a 1-to-1 cresol-xylene mixture, was ball-milled for 48 hours with the glass-frit refractory pigment. With the use of this resin the maximum pigment loading, with the proper retention of flexibility, was 50% pigment, 50% polyester resin solids.

Refractories

During the glass-fusing process, the refractory functions as a wire separator and partially dissolves into the glass to raise its fusion temperature. Among the many refractory materials investigated, silica (240 mesh) was selected as the best film former with the glass frit after the organic resin had been burned out and the glass fused.

Glass Frits

Since the glass frit must be fused after winding the wire into equipment, the glass had to be designed to have a fusion temperature which would not be injurious to the rest of the equipment. The glass frit must also retain sufficient insulation resistance and cut-through resistance at the operating temperature of the equipment. A minimum of glass frit which gave the proper binding action was required to eliminate cut-through during the glass-fusion operation. This was determined at 40% glass frit, 60% refractory. The first glass frit used was a 200-mesh lead borate glass with a 475 C (degrees centigrade) sag point. A formulation containing 50% polyester resin, 30% silica, and 20% glass frit was coated on no. 16 silver wire at 22 feet per minute and baked at 460 C in a 15-foot tower by the conventional dip-coating method.

Table I. Insulation Resistance, Megohms

Temperature, C	Trial 1	Trial 2
300	15,000	3,600
350	8,000	520
400	12,200	200
450	220	0
500	0	

The enameled wire had the following properties:

Before burnout:
- Build, mils, 3.0.
- Quick snap, O.K.
- Three-*X* mandrel, O.K.
- Repeated abrasion, 20, 30, 42 (General Electric Company repeated abrasion machine; 740-gram load; 16-mil diameter; knife edge).

After burnout:
- Build, mils, 1.2.

Table II. Enamel Properties

Properties	Conductor	
	Silver	Nickel-Plated Copper
Before Burnout		
Build, mils	3.2	3.8
Quick snap Three-*X* mandrel	O.K.	O.K.
Abrasion	30 25 27	40 30 35
After Burnout		
Adhesion	good	fair
Electric Strength	660 880 1,000	560 600 320
Cut-through temperature, C	678	710
Insulation resistance*	250	2,500

* At 550 C, determined on dielectric twisted pairs at 100 volts, megohms.

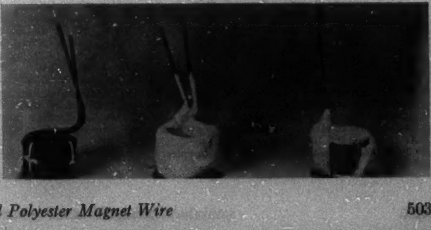


Fig. 1. Parallel-wound test coils

Paper 59-1295, recommended by the AIEE Electrical Insulation Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE–National Electrical Manufacturers Association Second National Conference on Application of Electrical Insulation, Washington, D.C., December 8–10, 1959. Manuscript submitted September 14, 1959; made available for printing January 29, 1960.

F. A. SATTLER and L. C. SCALA are with the Westinghouse Electric Corporation Research Laboratories, Pittsburgh, Pa.

The barium-modified glass frit used in the formulation described was developed in the Westinghouse Research Laboratories by Dr. C. Hirayama. The cement for the coils illustrated in Fig. 1 was developed by Dr. P. A. Tierney. D. K. McIlvaine supervised the winding of the transformers illustrated in Figs. 3 and 4.

Table III. Insulation Resistance, Megohms*

Degrees, C	Conductor No. 16 Silver	 No. 14 Nickel-Plated Copper
500	110	110
550	26	22
600	2.7	2.7
Breakdown†	300	370

* Readings taken after 10 minutes at given temperatures; coils burned out at 700 C before measuring.
† At 540 C, volts.

Appearance: smooth, hard, brittle, good adhesion.
Electric strength: 500 volts, 480 volts (wire-to-wire breakdown, standard National Electrical Manufacturers Association twist made and burned out at 650 C).
Cut-through temperature: 460 C, 440 C (500-gram load on crossed wires, temperature rise 5 C per minute from 100 C).

The burnout procedure was as follows:

	Temperature, C	
1/2 hour	350 450	with air flow
15 minutes	550 650	without air flow

Insulation resistance of the burned-out enamel was measured on standard dielectric twisted pairs using a megohm bridge at 100 volts; see Table I.

For the proper cut-through temperature and insulation resistance at 540 C, a 575 C sag point barium-modified glass frit was developed in the laboratories. This glass was incorporated into a formulation containing 50% polyester resin, 30% silica, and 20% glass frit. Coated on no. 16 American Wire Gage silver and nickel-plated copper, this enamel had the properties which are shown in Table II.

The burnout procedure was as follows:

	Temperature, C	
1/2 hour	350 450	with air flow
15 minutes	550 650 700	without air flow

Further insulation resistance and breakdown measurements were obtained on this enameled wire with parallel-wound coils, six double turns on a 3/4-inch mandrel tied with glass cord to force contact between turns; see Fig. 1. The coils were potted in a cement developed in the laboratories; it consisted of a magnesium aluminum silicate refractory bonded by a lead borate glass. The results are shown in Table III.

Parallel-wound coils of various conductors coated with the enamel just described were aged at 540 C with periodic heat cycling and insulation resistance measurements; no appreciable decrease in insulation resistance was present after 1,100 hours at 540 C; see Table IV and Fig. 2. The insulation resistance of the enameled nickel-plated copper wires, which was practically constant throughout most of the test period, increased noticeably after the seven consecutive heat cycles from room temperature to 540 C. No explanation for this phenomenon can be offered at this time. Thermal aging and cycling did not appear to have an appreciable effect on breakdown voltages at 540 C.

Metal Additives for Improved Adhesion after Burnout

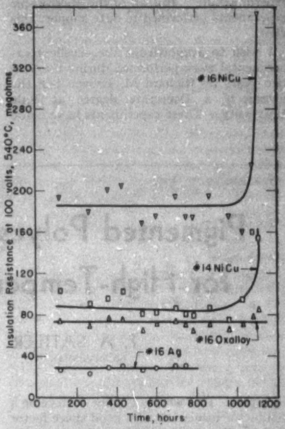


Fig. 2. Insulation resistance in megohms, 540 C versus aging time in hours at 54o C

Addition of soluble metal chelates to the enamel was studied in order to improve the vibration resistance of the residual inorganic coating after burnout. The best adhesion between the inorganic coating and the underlying conductor, particularly nickel-plated copper, was obtained when a mixture of cobalt and manganese was added to the formulation in a soluble form. The cobalt-manganese modified enameled wire gave the following properties on no. 16 nickel-plated copper wire:

Before burnout

- Build, mils, 3.2
- Quick snap, O.K.
- Three-*X* mandrel, O.K.
- Abrasion, 27, 33, 38.

After burnout

- Adhesion very good.
- Electric strength, R.T. (twisted pairs): 500 volts.
- Insulation resistance at 550 C (parallel-wound potted coil): 45 megohms.
- Electric strength at 550 C (parallel-wound potted coil): 360 volts.

Transformers were wound with the

Table IV. Insulation Resistance as Function of Time at 540 C and Voltage Breakdown at 540 C of Cement-Potted Coils, Megohms

Sample No.	Type of Conductor	Build Before Burnout, Mils	Insulation Resistance in Megohms at 100 Volts, at Various Times in Hours 120	 432*	 530*	 600*	 696	 745†	 782	 960*	 1,030*	 1,030‡	 1,080§	 1,110	Voltage Breakdown, Volts‖
1	No. 14 Ni Cu	5	170	145	100	110	115	110	108	100	95	80	170	160	240
2	No. 14 Ni Cu	5	56	71	62	55	60	52	50	50	42	110	150	150	260
3	Oxalloy	3.5	90	82	64	76	74	76	75	74	70	45	73	83	280
4	Oxalloy	3.5	57	74	60	65	74	64	54	55	53	100	85	85	260
5	No. 16 Ag	2.6	36	35	33	34	40	40	**						
6	No. 16 Ag	2.6	17	24	20	22	24	24	**						
7	No. 16 Ni Cu	4.1	90	120	100	93	97	95	85	90	83	130	470	470	400
8	No. 16 Ni Cu	4.1	300	290	240	260	300	260	260	240	230	320	280	250	360

* The furnace containing the samples was cooled to 100 C, then heated back to 540 C.
† The furnace containing the samples was cooled to room temperature, then heated back to 540 C.
‡ Readings taken after two consecutive cycles to 100 C from 540 C.
§ Readings taken after seven consecutive cycles to room temperature from 540 C.
‖ At 540 C, after 1,100 hours at 540 C.
** Tests discontinued for observation of samples.
Ni = nickel; Cu = copper; Ag = silver.

Fig. 3. Unpotted power transformers wound with metal-modified pigmented polyester enameled wire

metal-modified pigmented polyester enameled wire at the Company's Lima Works; see Figs. 3 and 4. They passed preliminary tests at 600 F (degrees Fahrenheit) ambient after potting in an inorganic cement.

The tests consisted of potting and burning out three power transformers wound with no. 18 enameled nickel-plated copper wire in the primary and no. 14 nickel-plated copper in the secondary. These were operated at room temperature and full load to determine the current, voltage, and reactance. The temperature rise of the core was approximately 89 F. The units were then given a thermal shock test by cycling from −65 to 600 F while being energized so as to draw full load current of 50 amperes on the secondary. Five additional cycles from −65 to 690 F were given the unenergized transformers, then a final cycle from −65 to 600 F while energized. The units were then vibration tested, the conductor insulation still remaining unaltered electrically.

Fig. 4. Power transformer potted and cured at 700 C

Conclusions

A conductor insulation containing a cobalt-manganese modified flexible polyester binder pigmented with a barium-modified lead borate glass frit and a silica refractory has been developed. Transformers wound with this enameled wire and potted in an inorganic cement have shown satisfactory performance at 600 F ambient in preliminary tests.

Theory and Operation of a Cross-Field Magnetic Modulator

D. W. BOENSEL
NONMEMBER AIEE

Synopsis: The theory of operation and practical limitations of a toroidal cross-field magnetic modulator are discussed. Approximate relationships are derived for the amplitude of the fundamental component of output voltage, conversion power gain, and output impedance. Experiments on devices of this kind using two types of ferrite core material are described. The agreement between experimental results and analytical calculations is satisfactory enough to warrant the use of the derived relations in design procedures. Among other things, it has been shown that the conversion power gain is directly proportional to excitation frequency and the number of turns on the input winding, and is independent of the number of turns on the output winding. Experimental results indicate that equivalent offset currents and voltages of 1 μa (microampere) and 50 μv (microvolts) respectively, and power gains of from 20 to 30 db (decibels) can be readily obtained.

THE PROBLEM of converting low-level direct currents or direct voltages to equivalent a-c information is unquestionably one of the most widely discussed problems in the instrumentation field. Common modulation techniques use electronic, magnetic, and mechanical devices to accomplish the shift of information from $\omega=0$ to some nonzero portion of the frequency spectrum.

The common magnetic frequency shifter is the conventional second-harmonic modulator, a device which has been adequately discussed in the literature.[1–3] A different approach to the problem using another magnetic device has been suggested by Heartz and Bueltemann.[4] This new approach employs a ferromagnetic medium under the influence of perpendicularly superimposed magnetic fields. The principal advantage of such a scheme is the elimination of inductive coupling between the excitation and output, which, of course, results in a converter exhibiting extremely low zero offset.

The purpose of this report is to describe the results of an investigation pertaining to a cross-field device of this nature. The results appear to be encouraging in that such a device can be readily constructed to have no more zero offset than the conventional second-harmonic modulator and furthermore, to have appreciable conversion power gain.

Theory of Operation

Although there are many core forms which can conceivably be used to permit winding of coils which establish geometric-

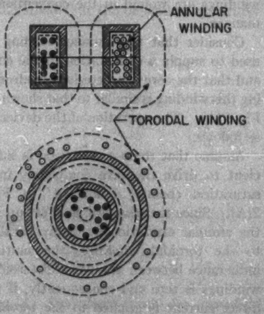


Fig. 1. Cross-field core

Paper 60-788, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting, Atlantic City, N. J., June 19–24, 1960. Manuscript submitted November 3, 1958; made available for printing April 5, 1960.

D. W. Boensel, formerly with the Jet Propulsion Laboratory of the California Institute of Technology, Pasadena, Calif., is now with the Space Electronics Corporation, Glendale, Calif.

This paper presents the results of one phase of research carried out at the Jet Propulsion Laboratory, California Institute of Technology, under Contract No. *DA-04-495-Ord 18*, sponsored by the U. S. Department of the Army, Ordnance Corps.

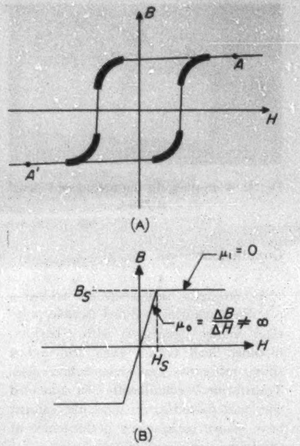


Fig. 2 (left). Annular core characteristic
A—Actual
B—Assumed

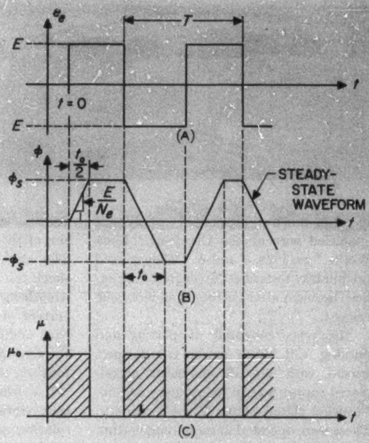


Fig. 3 (right). Waveforms pertinent to core cycling

ally orthogonal magnetic fields, the one chosen for this investigation consisted of a ferrite toroid having both annular and toroidal windings. Knowledge of the shape of the device is not particularly necessary for an explanation of its operation; however, since the experimental model took the abovementioned form, it will be expedient to explain the operation on that basis.

Preliminary Considerations

A sketch of the device is shown in Fig. 1. The core is made up of two ferrite toroids which are impact-ground to leave space for the annular winding. The surfaces of the toroids which are in intimate contact are lap-ground to reduce the air gap in the magnetic path surrounding the annular winding (henceforth referred to as the annular path). The cores are then placed securely together and toroidal windings are wound in the conventional manner.

Consider that the annular winding is used to supply a-c excitation to the core and that the core characteristic confronting this winding is similar to that shown in Fig. 2(A). The operation of the device is as follows.

Assume that the a-c excitation is sufficient to drive the core completely into saturation (points A and A' in Fig. 2(A). Since the magnetic field set up by the annular excitation is ideally parallel to the toroidal windings, the mutual inductance between annular and toroidal windings is zero and consequently, if no direct current is applied to the toroidal winding, no voltage is induced in the toroidal winding by the alternating annular current.

Based on its geometry, turns, and the core permeability, the toroidal winding has a certain inductance L_t, so that a voltage across it is of the form

$$e_t=\frac{d}{dt}(L_t i_t) \qquad (1)$$

where i_t is the current impressed on the winding by an external current source.

Each time the core is driven into and out of saturation, the darkened regions of the core characteristic in Fig. 2(A), the core permeability experiences a radical change with time. Since L_t is proportional to $\mu = d\mathbf{B}/dH$, for i_t equal to a constant current I, it is found that

$$\begin{aligned} e_t &= \frac{d}{dt}(L_t I) && (2)\\ &= IL_t'\left(\frac{d\mu}{dt}\right) \end{aligned}$$

where $L_t=\mu L_t'$. Equation 2 assumes that I is so small that it has negligible effect on the annular core characteristics, i.e., $\mu \neq f(I)$.

This cursory examination indicates the following points:

1. The magnitude of the output voltage is directly proportional to I.
2. This magnitude is proportional to the time rate of change of permeability and consequently to the excitation frequency (for a given excitation waveshape) and to the excitation amplitude (for a given frequency).
3. The fundamental frequency of the output voltage is twice that of the excitation since $d\mu/dt$ undergoes 2 cycles for every traversal of the core characteristic by the excitation.

Basic Relationships

The following discussion describes an approximate analysis which can be used to determine the output voltage, output impedance, and conversion power gain which can be expected from a given device.

Assume that the annular core characteristic has the piecewise linear form shown in Fig. 2(B), and that the excitation voltage is of the form shown in Fig. 3(A). The corresponding flux behavior shown in Fig. 3(B) follows, since, during the nonsaturated period

$$\phi=\frac{1}{N_e}\int e_e\,dt$$

and thus

$$t_0=\frac{2\phi_s N_e}{E} \qquad (3)$$

where t_0 is the time during which the core is out of saturation and N_e is the number of annular turns. The permeability time variation corresponding to this flux behavior is therefore as shown in Fig. 3(C).

Reference to equation 2 indicates that $d\mu/dt$ is the quantity which determines the alternating output of the cross-field modulator. Because of the nature of the approximation made for the **B-H** curve (the slope of **B** versus **H** is a discontinuous function), it follows that $d\mu/dt$ will always consist of an impulse train, regardless of the waveshape and amplitude of the excitation (as long as it periodically saturates the core). If the frequency spectrum of this impulse train is cal-

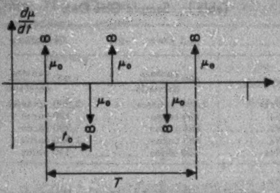


Fig. 4 (left). Impulse train in $d\mu/dt$

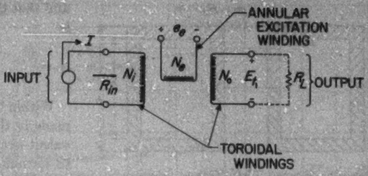


Fig. 5 (right). Circuit pertaining to calculation of conversion power gain

culated, and the magnitude of the fundamental component of the spectrum is determined, the result should give a reasonable approximation to the output from such a device using a square-loop core material. (Such an impulse train is, of course, impossible to obtain in practice. However, since a series approximation of the **B-H** curve would yield unwieldy results, the approximation given above was developed.) The impulse train of $d\mu/dt$ is shown in Fig. 4.

The coefficients in the $\mu(t)$ series are given by

$$c_k=\frac{4\mu_0t_0}{T}\left(\frac{\sin\left(2\pi k\frac{t_0}{T}\right)}{2\pi k\frac{t_0}{T}}\right) \tag{4}$$

Thus, in the $d\mu/dt$ series, the coefficients are given by

$$c_k'=\omega c_k=\frac{2\pi k}{T}c_k=\frac{8\mu_0}{T}\left[\sin\left(2\pi k\frac{t_0}{T}\right)\right] \tag{5}$$

and the amplitude of the fundamental component is

$$c_1'=\frac{8\mu_0}{T}\left[\sin\left(2\pi\frac{t_0}{T}\right)\right] \tag{6}$$

The fundamental output amplitude is, from equations 2 and 6,

$$|E_{t_1}|=\left|\frac{8\mu_0L_t'}{T}\left[\sin\left(2\pi\frac{t_0}{T}\right)\right]I\right| \tag{7}$$

The maximum value of this amplitude occurs when $t_0=T/4$ and is given by

$$|E_{t_1}'|_{\max}=\left[\frac{8\mu_0L_t'}{T}\right]|I| \tag{8}$$

For the purpose of obtaining an approximate expression for the conversion power gain of the device, it is expedient to investigate equation 7 more thoroughly. Toward this end equation 7 may be rewritten:

$$|E_t|_1=Z_{\text{eff}}|I| \tag{9}$$

where

$$Z_{\text{eff}}=|\omega_{t_1}L_{\text{eff}}|$$

$$\omega_{t_1}=\frac{2\pi}{\tau}=\frac{4\pi}{T}$$

=fundamental output frequency

$$\tau=\frac{T}{2}=\text{period of output voltage}$$

$$L_{\text{eff}}=\frac{2}{\pi}\left[\sin\left(2\pi\frac{t_0}{T}\right)\mu_0L_t'\right] \tag{10}$$

The implications of equation 9 may be interpreted by considering the circuit shown in Fig. 5. Since both input and output windings are toroidal, the flux introduced by I at the input is transferred undiminished to the output winding. In other words, the flux in the toroidal path is the same flux which would be initiated in the core if a current of $(N_t/N_0)I$ were made to flow in the output winding. (One toroidal winding excited with a control current could just as well be used, coupled to R_L through a coupling capacitor.) Thus, the amplitude of the fundamental component of output voltage (E_{t_1}) is

$$|E_{t_1}|=Z_{\text{eff}}\left|\frac{N_t}{N_o}I\right| \tag{11}$$

The effective output impedance of the device at the fundamental output frequency is

$$Z_o=Z_{\text{eff}} \tag{12}$$

so that the maximum power which can be supplied to a load, R_L, occurs when

$$R_L=Z_o=Z_{\text{eff}}=|\omega_{t_1}L_{\text{eff}}| \tag{13}$$

(This is for an untuned or purely resistive load.)

Under these conditions, the a-c load power is

$$P_o=\frac{|E_{t_1}|^2}{8\mathrm{R}_L}=\frac{Z_{\text{eff}}\left|\frac{N_t}{N_o}I\right|^2}{8} \tag{14}$$

and the input d-c power is $P_{\text{in}}=I^2R_{\text{in}}$ so that the conversion power gain of the device is

$$\mathrm{PG}_c=\frac{P_o}{P_{\text{in}}}=\frac{Z_{\text{eff}}}{8R_{\text{in}}}\left(\frac{N_t}{N_o}\right)^2$$
$$=\frac{\omega_{t_1}}{4\pi R_{\text{in}}}\left[\sin\left(2\pi\frac{t_0}{T}\right)\right]\mu_0L_t'\left(\frac{N_t}{N_o}\right)^2 \tag{15}$$

where PG_c=conversion power gain.

In the practical meter-kilogram-second system, L_t' is given by

$$L_t'=\frac{(4\pi\times10^{-7})N_o^2A_t}{l_t}$$

where A_t=cross-sectional area of the toroidal path in square meters, and l_t= mean toroidal path length in meters.

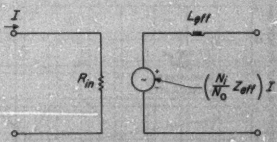
Equation 15 may now be written

$$\mathrm{PG}_c=\frac{\omega_{t_1}\mu_0A_tN_t^2\times10^{-7}\left[\sin\left(2\pi\frac{t_0}{T}\right)\right]}{l_TR_{\text{in}}} \tag{16}$$

(In this formulation μ_0=gausses/oersteds)

It is interesting to observe that in this analysis the conversion power gain is independent of the number of turns on the output winding. Obviously this does not imply that N_o may go to zero, for, if this were the case, the output voltage and consequently the load power would likewise become zero. It does indicate, however, that one may choose N_o to provide any desired output impedance level and not cause any deterioration in power gain. It is also worth while to note that although equation 16 states that the power gain is proportional to N_t^2, for a given wire size, R_{in} is also proportional to N_t, so that the power gain is actually proportional to N_t rather than N_t^2. If a fixed available winding area is assumed, however, the power gain is independent of N_t, as would be expected.

Based on the relations derived in equations 7 and 12, an equivalent circuit depicting the operation of the cross-field modulator can be developed. Such a circuit is shown in Fig. 6. It should be understood that the dependent generator



$$L_{eff}=\frac{(8\times10^{-7})N_o^2\mu_0A_t}{l_t}\left[\sin\left(2\pi\frac{t_0}{T}\right)\right]$$

$$Z_{eff}=\frac{(32\pi\times10^{-7})N_o^2A_t\mu_0}{Tl_t}\left[\sin\left(2\pi\frac{t_0}{T}\right)\right]$$

Fig. 6. Equivalent circuit for cross-field modulator

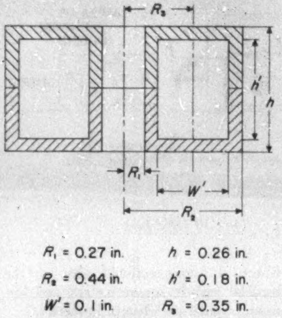


Fig. 7. Experimental core dimensions

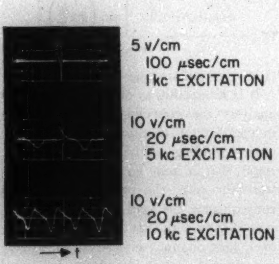


Fig. 8. Representative output waveforms

in this equivalent circuit supplies an alternating voltage having a fundamental component at twice the frequency of the excitation.

With this equivalent circuit, it is a simple matter to show that the open-circuit voltage gain A_v is

$$A_v = \frac{N_t}{N_o}\frac{Z_{eff}}{R_{in}}$$

$$= \frac{(32\pi\times 10^{-7})N_o N_t A_t \mu_0}{T l_t R_{in}}\left[\sin\left(2\pi\frac{t_0}{T}\right)\right] \tag{17}$$

and that the short-circuit current gain A_t is

$$A_t = \frac{N_t}{N_o} \tag{18}$$

Within the limits of the approximations made in the foregoing analysis, the equivalent circuit shown in Fig. 6 can be used to formulate all of the circuit properties necessary for the design of a toroidal cross-field modulator.

Experimental Results

As mentioned previously, the form of core configuration used in experiments is that shown in Fig. 1. Physical dimensions of the cores are given in Fig. 7. The following table lists the corresponding annular and toroidal winding magnetic cross sections and path lengths.

	Annular	Toroidal
Cross-section area, cm²	0.39	0.17
Mean path length, cm	1.8	5.7

Two types, type *S-1* and type *H*, of General Ceramics Corporation ferrite core material were employed in separate experiments. Pertinent core characteristics as well as winding data for the two samples are given in Table I. The values of μ_0 (not to be confused with initial permeability) indicated for these materials were found by extrapolating an average permeability from the measured core characteristics for excitation amplitudes approximating those used experimentally. These values do not correspond to those given by the manufacturer, probably because of the air gap which is present in the annular path. Unfortunately it was impossible to measure the air gap, so experimental **B-H** loops were necessarily employed in this calculation.

The experiments consisted in simply measuring the amplitude of the fundamental component of output voltage as a function of excitation frequency and input current. The load resistance necessary to reduce the amplitude of this voltage to one half its open-circuit value was also measured to obtain the output impedance as a function of frequency.

Table I. Sample Core Data

	S-1 Core	H Core
μ_0	800 $\frac{\text{gausses}}{\text{oersteds}}$	1,000 $\frac{\text{gausses}}{\text{oersteds}}$
B_{sat}	1,500 gausses	2,800 gausses
N_e	125 turns	100 turns
N_t	885 turns	1,500 turns
N_o	885 turns	1,500 turns
R_{in}	50 ohms	150 ohms

Representative output voltage waveforms under open-circuit conditions on the *S-1* core are shown in Fig. 8. The waveforms are shown for excitation frequencies of 1, 5, and 10 kc and an input current of 400 μa. It is interesting to observe that the output voltage at low frequencies is a reasonable facsimile of the impulse waveform predicted in Fig. 4. It should be noted, however, that as the frequency is increased, the lengths of the pulse generated when the core enters and leaves saturation become an appreciable portion of the total period of the output voltage. This is a result of a softening of the knee of the **B-H** curve at higher frequencies, as well as in increase in the pulse-stretching effect of distributed winding capacitance. At any rate, it is obvious that when this region of operation is reached, the analysis presented in the previous section can be expected to give only very approximate results.

The following paragraphs describe the comparison of experimental and analytical results on output voltage amplitude, output impedance, conversion power gain, and linearity.

OUTPUT VOLTAGE AMPLITUDE

The fundamental output voltage amplitude was measured with a Donner wave

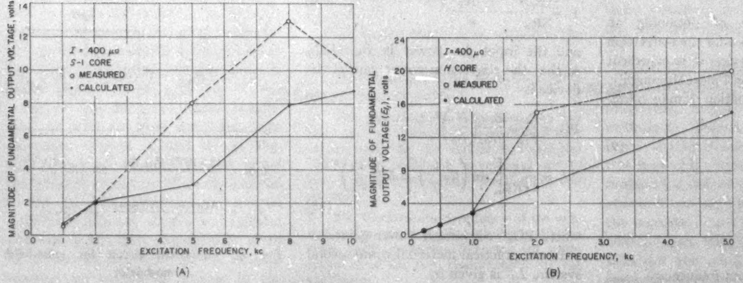
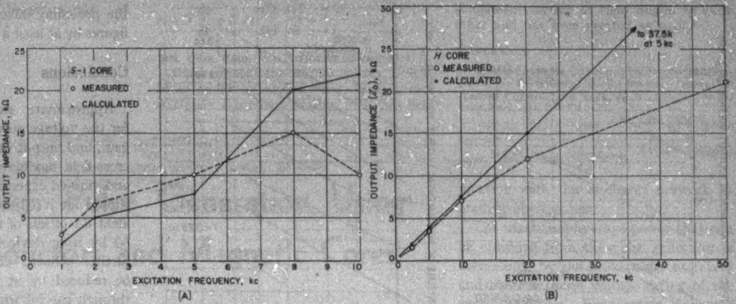


Fig. 9. Output amplitude versus excitation frequency

Fig. 10. Output impedance versus excitation frequency



analyzer. The results of these measurements, together with calculated values, are shown in Figs. 9(A) and 9(B). Some difficulty was encountered with the **H** core in that an extra unused toroidal winding caused the output to "ring." To damp out these oscillations, it was necessary to load the winding with a resistance approximately equal to its output impedance. Therefore, the measured voltage amplitudes were multiplied by a factor of 2, and 6 db was added to the measured conversion power figures to account for the loading on the winding.

In general, agreement of the results for frequencies below 2 kc is quite satisfactory. Even for frequencies up to 10 kc, the predicted values are off, in the worst situation, by only a factor of two. As predicted, there is a tendency for the output voltage to increase as a linear function of frequency. The reduction in this tendency at higher frequencies is obviously a result of the change in waveshape from an impulse train to a quasi-sinusoid which was observed in Fig. 8. To clarify this point, it should be observed that, whereas the amplitude of all components in an impulse train increases linearly with frequency, the amplitude in a pulse train made up of pulses having nonzero width is independent of the absolute magnitude of the fundamental frequency. Thus, as the output pulses stretch (relative to the total period) the tendency for a linear increase in amplitude with frequency diminishes.

It should be mentioned that the predicted curve in Fig. 9(A), as well as that in Fig. 10(A), is not a smooth straight line because of the variation of the term sin $2\pi t_0/T$ with frequency at constant excitation voltage amplitude. The experiments relating to Figs. 9(B) and 10(B) were performed with changes in excitation amplitude in order to maintain this term constant.

Output Impedance

The results of impedance measurements are shown in Figs. 10(A) and (B). For both types of core material, the output impedance is initially a linear function of frequency, as predicted. As the excitation frequency increases, the **B-H** loop becomes less square, so that relatively more time is spent in a region of low permeability resulting in a smaller average value of effective output inductance. This obviously results in a relationship which is somewhat less than a linear function of frequency. Over and above this is the effect of distributed capacitance, which likewise serves to reduce the output impedance below predicted values at high frequencies.

This tendency is evident in both sets of experimental data; nevertheless, the maximum error between calculated and measured impedance values occurs at a high frequency and yet amounts to a factor of only 1.8; see the results for the **H** core in Fig. 10(B) at 5 kc.

Conversion Power Gain

For most applications, one would not design a cross-field modulator to provide a specified amount of conversion power gain. Usually the quantities of interest would be the levels of output and input impedances and the voltage output per unit current input. It is generally of only academic interest that all of these quantities uniquely determine the maximum available power gain for an untuned load.

The results of these experiments and analytical calculations are shown in Fig. 11.

Although the agreement of the overall experimental and theoretical computations is somewhat less than highly satisfactory, it is nevertheless worth while to note that the low-frequency results are

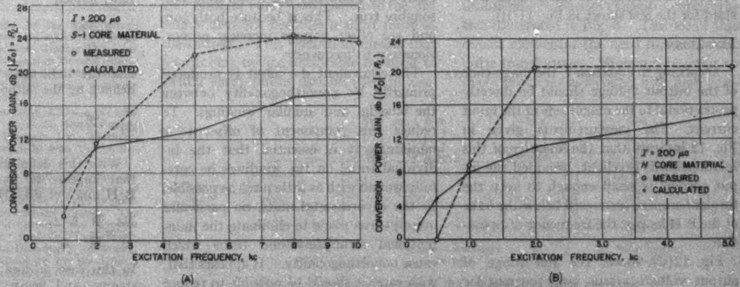


Fig. 11. Conversion power gain versus excitation frequency

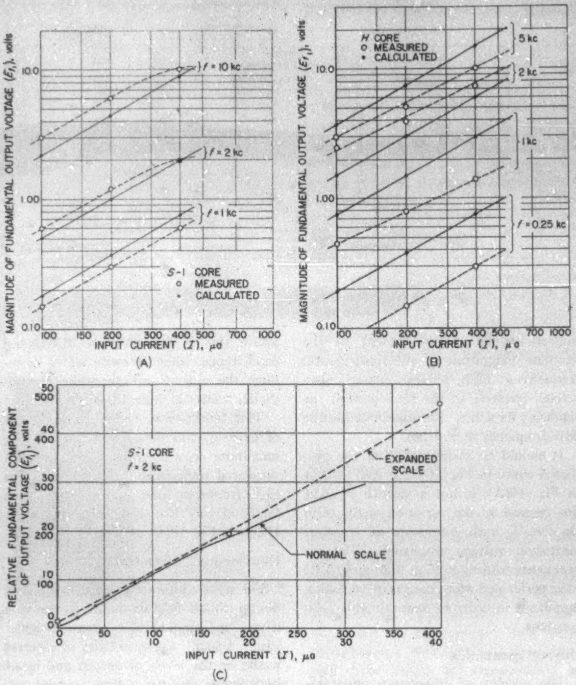


Fig. 12. Output voltage versus input current

A, B—Actual
C—Relative

usable for design procedures and those at the higher frequencies are adequate for approximate estimates.

The departure of the predicted curve of Fig. 11(A) from a smooth parabola is again caused by the variation of the sine term with frequency. As indicated previously, this term was maintained constant for the test shown in Fig. 11(B).

Linearity

According to equation 7, the magnitude of the output voltage should be directly proportional to the magnitude of the input current. The measurements given in Fig. 12 indicate that the linearity of this relationship is excellent, provided the input current is small enough to keep the toroidal ampere-turns well below the knee of the **B-H** loop at the frequency of operation.

Fig. 12(C) shows the full range of output voltage versus input current for the *S-1* core at 2 kc. The zero offset in this case is equivalent to approximately 1 μa at the input. Since the input resistance in this case is 50 ohms, the equivalent input voltage is 50 μv and the power is 5×10^{-11} watt. The expected result of equivalent input d-c offset remaining constant with frequency was found to be reasonably true. This is because both gain and a-c offset voltage increase proportionately to frequency.

The equivalent d-c offset is produced primarily by nonorthogonality between the toroidal and annular windings. To reduce this component of offset to a minimum, it is essential that the individual turns on the windings be carefully wound with as little pitch as possible. In the experimental units, no particular attempt was made to eliminate the unintentional scramble-wound turns which cause nonorthogonality. It appears that, with care, it should be possible to reduce the preceding offset voltage and current figures by at least a factor of ten.

Conclusions

Approximate analytical relationships for the voltage output, conversion power gain, and output impedance of a toroidal cross-field modulator have been derived and verified experimentally. The results show that it is possible to expect equivalent d-c offsets of the order of 1 μa and 50 μv from devices of this type. It is further believed that these figures can be reduced by at least a factor of ten through careful winding techniques.

From the standpoint of linearity, the device appears to have excellent possibilities as a d-c to a-c converter which can have power gains of the order of 20 to 30 db.

References

1. Magnetic-Amplifier Circuits (book), W. A. Geyger. McGraw-Hill Book Company, Inc., New York, N. Y., 1954, pp. 223–26.

2. Magnetic Amplifiers (book), H. F. Storm. John Wiley & Sons, Inc., New York, N. Y., 1955, pp. 402–03, 460–62.

3. Design Criteria for Low-Level Second-Harmonic Magnetic Modulators, E. J. Kletsky, *Technical Report no. 1, D.I.C. Project no. 6896*. Massachusetts Institute of Technology, Cambridge, Mass.

4. The Application of Perpendicularly Superposed Magnetic Fields, R. A. Heartz, H. Buelteman, Jr. *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 74, Nov. 1955, pp. 655–60.

◆

Discussion

F. J. Friedlaender (Purdue University, Lafayette, Ind.): The author presents an interesting analysis to explain the operation of a cross-field magnetic modulator. However, several questions arise with regard to some of the derivations and concepts on which the paper is based.

Equation 1 evidently expresses Faraday's Law, namely

$$e_t = \frac{d(N\phi)}{dt}$$

and the inductance L_t of equation 1 is defined by the author as

$$L_t = \frac{N\phi}{i_t}$$

With this definition, the permeability μ that appears in the expression of L_t is **B**/**H** and not $d\mathbf{B}/d\mathbf{H}$. In fact,

$$\frac{N\phi}{i_t} = \frac{\mathbf{B}}{\mathbf{H}} \frac{N^2 A_t}{l_t}$$

In this case, μ does not display a discontinuity, and hence the author's analysis and predictions no longer apply. (A differential permeability

$$\mu_{diff} = \frac{d\mathbf{B}}{d\mathbf{H}}$$

could be used; e.g.

$$e_t = N\frac{d\phi}{dt} = NA_t\frac{d\mathbf{B}}{d\mathbf{H}}\frac{d\mathbf{H}}{dt} = \frac{N^2A_t}{l_t}\mu_{diff}\frac{di}{dt}$$

but then equation 2 is not obtained.)

Furthermore, the author treats **B** and **H** as scalar quantities when, in fact, this is clearly a problem dealing with two field components acting at right angles to each other and has been treated as such.[1]

REFERENCE

1. FERROMAGNETIC COUPLING BETWEEN CROSSED COILS, V. F. Gianola, D. B. James. *Journal of Applied Physics*, New York, N. Y., vol. 27, no. 5, June 1956, pp. 608–09.

High-Power Pulse Generation Using Semiconductors and Magnetic Cores

EDWARD M. LASSITER
NONMEMBER AIEE

PAUL R. JOHANNESSEN
ASSOCIATE MEMBER AIEE

RICHARD H. SPENCER
ASSOCIATE MEMBER AIEE

Synopsis: A magnetic pulse generator using a controlled rectifier in the charging circuit is capable of generating high-power pulses and is suitable as a radar modulator. In the discharge network saturable inductors operate as high-power switching devices and, together with the capacitors, reduce the duration of the current pulse from a long duration at the charging circuit to a short duration at the output.

An experimental circuit demonstrates the theory. This circuit operates at a repetition rate of 5 kc and produces an average power output of 300 watts. The peak power and pulse width at the output are approximately 40 kw and 1.5 μsec (microseconds), respectively. The efficiency of the experimental circuit, not considering the bias circuit, is approximately 70%.

ADVANCES IN semiconductor devices have extended their use from the milliwatt power range to the kilowatt power range. This paper describes the application of one of the new semiconductor devices, the controlled rectifier, to radar power modulator circuits. The controlled rectifier switches power into a magnetic discharge circuit to accomplish this function. An example of the circuit configuration used is shown in Fig. 1. The circuit[1] is in some respects similar to previously developed circuits.[2,3]

The circuit operation can be explained as follows. Assume that saturable transformer P_1 is biased in negative saturation and a pulse from the trigger source turns the electronic switch on. In this manner the supply voltage E is applied as a step to the L-C (inductance capacitance)[2,3] network and the resulting voltage build-up across capacitor C_1 drives the flux in P_1 from negative to positive saturation. Saturable inductor P_1 saturates when the voltage on capacitor C_1 reaches a maximum. At this time the charging current is zero, the electronic switch opens, and capacitor C_1 discharges through the saturated inductance of P_1 to capacitor C_2. The voltage build-up on C_2 drives the flux in saturable inductor P_2, which was biased in negative saturation, from negative to positive saturation. When the voltage on capacitor C_2 reaches a maximum, saturable inductor P_2 saturates and C_2 discharges into the next section of the discharge network.

Each capacitor discharges more rapidly than it charges. In this way a relatively long pulse of energy switched into the discharge network becomes a short pulse at the output. This operation means that a switching device unsuitable by itself to switch short pulses can be used in conjunction with the discharge network to produce short pulses.

The maximum average power that can be obtained from the pulse generator is determined by the peak recurrent rating and the maximum voltage rating of the switching device. For example, if a power transistor such as the *2N-574* transistor is used, the maximum average power is approximately 50 watts. For applications such as radar modulators it is often desirable to have larger average power capabilities. The controlled rectifier offers a large increase in power handling capabilities of this type of pulse generator and this paper considers its use for such a purpose.

Circuit Configuration

Two general forms of the magnetic pulse generator exist: the shunt capacitor-series saturable inductor configuration, and the shunt saturable inductor-series capacitor configuration. An example of the shunt capacitor-series saturable inductor circuit is shown in Fig. 2. When the shunt saturable inductor-series capacitor circuit is used, the saturable inductor can be used as a combination switch and transformer. An example of this type of circuit is shown in Fig. 3.

A load, such as a magnetron in a radar application, requires a voltage in the order of several kilovolts. Thus, a pulse generator of the type shown in Fig. 2 requires a pulse transformer at the output to obtain the necessary step up in voltage. A more convenient and economic method of voltage step up is to use the shunt saturable inductor-series capacitor configuration illustrated in Fig. 3. The use of this circuit eliminates the need for an output pulse transformer. However, in order to avoid problems associated with high volt-

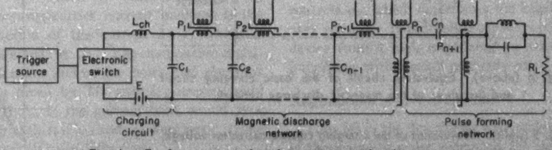


Fig. 1. Configuration used in the experimental pulse generator

---

Paper 60-870, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting, Atlantic City, N. J., June 19–24, 1960. Manuscript submitted June 18, 1959; made available for printing May 9, 1960.

EDWARD M. LASSITER is with Chance Vought Aircraft, Dallas, Texas. PAUL R. JOHANNESSEN is with Sylvania Electric Products, Inc., Waltham, Mass. RICHARD H. SPENCER is with United Research, Inc., Cambridge, Mass.

This paper is based on a thesis submitted by Edward Lassiter to the Massachusetts Institute of Technology in partial fulfillment of the requirements for the Master of Science degree. The work was made possible through the support extended the Electronic Systems Laboratory, formerly Servomechanisms Laboratory, of Massachusetts Institute of Technology by the United States Air Force under Contract *AF33-(616)-5489*, Task *50688*.

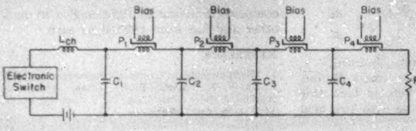


Fig. 2. Series saturable inductor circuit

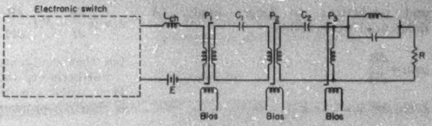


Fig. 3. Shunt saturable inductor circuit

age, it is desirable to locate most of the voltage step up in the last stage.

The best configuration as far as copper loss and load requirement are concerned is a combination of the two configurations. The shunt saturable inductor-series capacitor configuration is used in the last stage to eliminate the need for an external pulse transformer, the shunt capacitor-series saturable inductor configuration being used in the remaining stages to minimize the copper loss. A configuration of this type is used in the experimental model and is illustrated in Fig. 1.

ANALYSIS OF CIRCUIT

The charging circuit of Fig. 1 is redrawn for convenience in Fig. 4 with the first saturable inductor replaced by a switch in parallel with a constant current source. The two switches shown are assumed to operate periodically in the proper sequence as previously explained. The switch labeled S_e represents the electronic switch, and the switch labeled S_{p1} represents the first saturable inductor. The constant current source in parallel with switch S_{p1} represents the bias current, reflected in the terminals of saturable inductor P_1; magnetizing current is neglected.

It is convenient to separate the period of operation into two parts, the charging interval and the resetting interval. During the resetting interval switches S_e and S_{p1} are open and the constant bias current produces a linear voltage build-up on capacitor C_1 to reset saturable inductor P_1. It is assumed that the resetting interval has ended at $t=0$ so that capacitor C_1 is charged to a voltage $-\gamma$. At time $t=0$ switch S_e closes. The resulting voltage build-up on capacitor C_1 during the charging interval can be expressed as the sum of three voltages,

$$e_{C_1}(t)=e_1(t)+e_2(t)+e_3(t) \qquad 0<t<t_0 \qquad (1)$$

where $e_1(t)$ is the voltage build-up caused by the d-c supply, $e_2(t)$ is caused by the initial charge on capacitor C_1, and $e_3(t)$ is the voltage build-up caused by the constant bias current. These voltages can be expressed as follows:

$$e_1(t)=E(1-\cos\omega t) \qquad 0<t<t_0 \qquad (2)$$

$$e_2(t)=-\gamma\cos\omega t \qquad 0<t<t_0 \qquad (3)$$

$$e_3(t)=-\frac{I_b}{C\omega}\sin\omega t \qquad 0<t<t_0 \qquad (4)$$

where

$$\omega=\sqrt{\frac{1}{L_{ch}C_1}}$$

and the voltage on C_1 becomes

$$e_{C_1}(t)=E(1-\cos\omega t)-\gamma\cos\omega t-\frac{I_b}{C\omega}\sin\omega t \qquad 0<t<t_0 \qquad (5)$$

The total current flowing through the electronic switch and charging inductor during the charging interval can be expressed as

$$i_{ch}(t)=I_b-C_1\frac{de_{C_1}(t)}{dt} \qquad (6)$$

Differentiating and collecting terms yield

$$i_{ch}(t)=(E+\gamma)\sqrt{\frac{C_1}{L_{ch}}}\sin\omega t+I_b(1-\cos\omega t) \qquad 0<t<t_0 \qquad (7)$$

At the end of the charging interval, the charging current is zero and switch S_e opens. In this manner, the electronic switch is operated only when the current through it is zero and therefore little power is dissipated in it during the switching interval. If the time at which the current is zero is denoted as time $t=t_0$, then equation 7 yields

$$(E+\gamma)\sqrt{\frac{C_1}{L_{ch}}}\sin\omega t_0+I_b(1-\cos\omega t_0)=0 \qquad (8)$$

The waveforms of the current, capacitor voltage, and voltage across the switch S_e are shown in Fig. 5.

Another equation relating the variables of equation 8 may be obtained by considering the requirement for periodic operation of saturable inductor P_1. The volt-time integral across P_1 during the charging interval is the integral of the capacitor voltage and is given by the expression

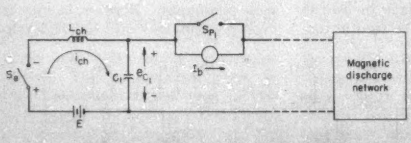


Fig. 4 (above). Equivalent circuit of the basic charging circuit and first stage of the magnetic discharge network

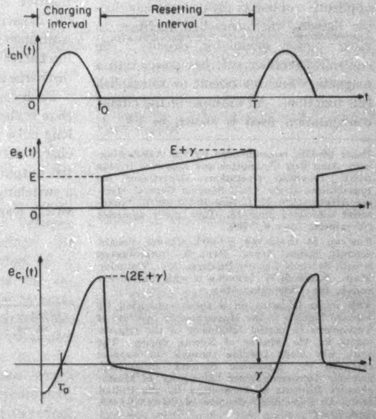


Fig. 5 (right). Waveforms of the charging current, capacitor voltage, and voltage across the switch in the charging circuit

$$\lambda_{ch}(t)=\int_0^{t_0} e_{C_1}(t)dt \qquad (9)$$

Substitution of equation 5 in equation 9, and integrating yields

$$\lambda_{ch}(t)=Et_0-\left(\frac{E+\gamma}{\omega}\right)\sin\omega t_0+\frac{I_b}{C_1\omega^2}(\cos\omega t_0-1) \qquad (10)$$

The voltage on capacitor C_1 during the resetting interval can be expressed as

$$e_{C_1}(t)=-\left(\frac{\gamma}{T-t_0}\right)(t-t_0) \qquad t_0<t<T \qquad (11)$$

The integral of the voltage across saturable inductor P_1 during the resetting interval is approximately equal to the integral of the voltage across capacitor C_1 and may be expressed as

$$\lambda_{rs}(t)=-\frac{\gamma}{T-t_0}\int_{t_0}^{T}(t-t_0)dt \qquad t_0<t<T \qquad (12)$$

It is assumed in the foregoing expression that the first loop of the magnetic discharge network is effectively decoupled from the remaining loops by the action of saturable inductor P_2, and that the impedance of P_2 is negligible compared to the other elements in the loop. After integration and simplification, equation 12 becomes

$$\lambda_{rs}(t)=-\frac{\gamma}{2}(T-t_0) \qquad (13)$$

Since the integral of the voltage across saturable inductor P_1 during one cycle of operation must be zero, the addition of equations 10 and 13 yields

$$Et_0-\left(\frac{E+\gamma}{\omega}\right)\sin\ \omega t_0+\frac{I_b}{C_1\omega^2}(\cos\omega t_0-1)-\frac{\gamma}{2}(T-t_0)=0 \qquad (14)$$

The following expression relates I_b to the previously defined parameters

$$I_b=\frac{\gamma C_1}{T-t_0} \qquad (15)$$

If equation 15 is substituted in equation 8, and the resulting equation is solved for γ, then

$$\gamma=-\frac{E\omega(T-t_0)\sin\omega t_0}{\omega(T-t_0)\sin\omega t_0+(1-\cos\omega t_0)} \qquad (16)$$

Substitution of this value of γ into equation 14 and simplifying the resulting expression yields

$$\omega t_0(1-\cos\omega t_0)+\sin\omega t_0\left[\frac{(\omega T)^2}{2}-\frac{(\omega t_0)^2}{2}\right]=0 \qquad (17)$$

This equation can be simplified if the following substitution is made:

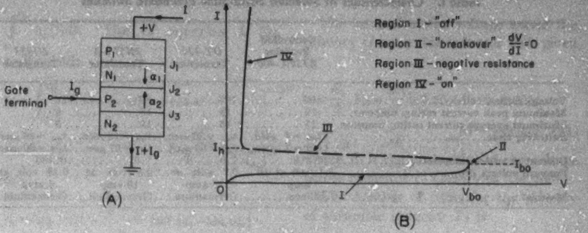


Fig. 6. P-n-p-n semiconductor

A—Schematic
B—Typical V-I characteristic

$$T=Kt_0 \qquad (18)$$

where K is the reciprocal of the duty ratio of the charging circuit. Substitution of equation 18 in equation 17 yields:

$$(1-\cos\omega t_0)+\frac{\omega t_0}{2}(K^2-1)\sin\omega t_0=0 \qquad (19)$$

By selecting appropriate values of K, equation 19 may be solved for ωt_0, which in turn can be used to obtain expressions for the various voltages and current in the charging circuit.

Controlled Rectifier as an Electronic Switch

The controlled rectifier is a p-n-p-n silicon semiconductor. It is a 3-terminal device and, in many respects, is similar in behavior to the conventional thyratron. A schematic of the p-n-p-n seniconductor device is shown in Fig. 6(A).

The junctions J_1 and J_3 are called the emitter junctions and J_2 is called the collector junction. If the gate terminal is left open-circuited, the device may be considered as a 2-terminal device of the type investigated by Moll et al.[4] A typical characteristic of such a device is shown in Fig. 6(B). If V is positive as shown in Fig. 6(A), the collector junction J_2 is reverse biased, and the emitters J_1 and J_3 are slightly forward biased. The current I which flows through the junctions is given by

$$I=IM_p\alpha_{1n}+IM_n\alpha_{3n}+I_{co} \qquad (20)$$

where α_{1n} is the fraction of the current at J_1, which is collected at J_2 as minority carrier current and α_{3n} is the fraction of the current at J_3 which is collected at J_2. The avalanche multiplication factors for holes and electrons are given by M_p and M_n respectively,[5] and I_{co} is the current that would flow through J_2 if α_{1n} and α_{3n} were zero. If the total alpha is defined as

$$\alpha_T=M_p\alpha_{1n}+M_n\alpha_{3n} \qquad (21)$$

then solution of equation 20 for I yields

$$I=\frac{I_{co}}{1-\alpha_T} \qquad (22)$$

The four principal regions of the V-I characteristic in Fig. 6(B) are explained as follows:[4]

1. At low currents, α_T is less than 1 and J_2 behaves essentially like a classical p-n junction in reverse bias. This is termed the "off" condition (high impedance region).
2. As the applied voltage is increased, avalanche multiplication at J_2 increases to a point where $\alpha_T=1$ and J_2 breaks down. This is called the "breakover" region, defined by $dV/dI=0$, where the voltage and current in 2-terminal operations have the values V_{bo} and I_{bo} respectively.
3. The device characteristic then traverses a region of negative resistance as shown by the dotted line in Fig. 6(B).
4. The low impedance region is reached when $\alpha_1+\alpha_2\geq 1$; at these low voltages M_n and M_p are essentially unity. The relation expressed in equation 22 is not valid in this region, and the device behaves as a resistor of low value. In the "on" region a certain minimum current is required to maintain the device in the low impedance state. This minimum current is designated as the holding current, I_h.

The above description relates to the controlled rectifier when no gate current is applied. If gate current is applied, the breakover voltage can be varied. The gate controls α_T, which determines the breakover voltage. It is possible to switch the device from a high impedance to a low impedance with a small gate current.

REQUIREMENTS OF THE SWITCH

Thus far in the description and analysis of the magnetic pulse generator, the elec-

Table I. Characteristics of Selected Solid-State Electronic Switches

	Controlled Rectifier *ZJ39A-400*	*DT-100* Transistor	*2N574-A* Transistor	*2N458* Transistor
Voltage rating, volts	400	100	80	80
Maximum peak current rating, amperes	20	13	15	5
Maximum average current rating, amperes	15	13	15	5
Switching time	$t_{on}=1$ μsec $t_{off}=3$ μsec	$t_{on}=30$ μsec $t_{off}=60$ μsec	$t_{on}=15$ μsec $t_{off}=27$ μsec	$t_{on}=26$ μsec $t_{off}=80$ μsec
Current to switch, amperes	0.010	1.0	1.5	0.200
Forward voltage drop	1 volt at 10 amp	0.7 volt at 12 amp	0.5 volt at 10 amp	0.24 volt at 5 amp
Material	Silicon	Germanium	Germanium	Germanium

tronic switch has been assumed to have ideal properties. In order to determine the suitability of the controlled rectifier as the switch, the pertinent properties of possible solid-state switches are listed in Table I.

Present-day power transistors set an upper limit on the power available from the magnetic pulse generators using a single transistor in the charging circuit. The emitter-collector voltage ratings (60 to 80 volts) and maximum collector current ratings (12 to 15 amperes) have restricted the output to a medium power range (20 to 50 watts). Since the output power depends upon the d-c supply voltage and current in the charging circuit, any increase in power must result from increasing these values. The controlled rectifier permits the use of voltages up to 400 volts and peak recurrent currents up to 30 amperes.

Another desirable property of the controlled rectifier is its ease of control. The controlled rectifier can be turned on by a small pulse of gate current, a typical requirement being 10 milliamperes at 1.5 volts. Besides the ease with which the controlled rectifier is switched into conduction, it automatically turns itself off when the current in the anode circuit is zero. This property is particularly well suited for the present application because this is the exact time that the switch should open. Whereas complex circuitry may be required to switch a transistor off when the charging current is zero, it is accomplished automatically with the controlled rectifier.

Typical turn-on and turn-off times for the controlled rectifier are 1 and 3 μsec respectively. These values are better than the typical transistor values of 12 to 40 μsec for turn-on time and 25 to 80 μsec for turn-off time. The only switching property in which the controlled rectifier is not superior to other presently available semiconductors is the voltage drop during conduction. The forward voltage drop across the controlled rectifier is typically 1 volt at 10 amperes. Typical values of transistor forward voltage drop in the corresponding current range are 0.4 to 0.7 volt.

Experimental Pulse Generator

The experimental circuit was designed to meet the following requirements:

1. average output power: 250 watts
2. pulse repetition frequency: 5 kc
3. output pulse duration: 1 μsec

The first step is the design of the charging circuit. The maximum peak recurrent rating of the controlled rectifier is 30 amperes. If the efficiency of the generator is estimated at 57%, from previous circuits, and a d-c supply voltage of 80 volts is chosen, then a first approximation to the input energy per pulse is 250/(5,000) (0.57) or 0.088 joule. The time required to transfer this energy at 80 volts with a half-sinusoid of current is $(0.088)\pi/(2)\ (30)\ (80)=58$ μsec. Thus K of equation 18 becomes $200/58=3.5$ and substituting this value into equation 19 yields

$$1-\cos\omega t_0+5.6\omega t_0\sin\omega t_0=0 \tag{23}$$

The value of ωt_0 which satisfies equation 23 is

$$\omega t_0=186.3° \tag{24}$$

Since $T=Kt_0$ and $\omega=1/\sqrt{L_{ch}C_1}$, an equation involving the charging inductance L_{ch} and capacitor C_1 can be written

$$\sqrt{L_{ch}C_1}=\frac{1}{\omega}=\frac{T}{K\theta} \tag{25}$$

where θ is the radian value of ωt_0. Substituting $K=3.5$, $\theta=3.24$ radians, and $T=200$ μsec yields

$$\sqrt{L_{ch}C_1}=\frac{T}{K\theta}=\frac{200\times10^{-6}}{3.5\left(\dfrac{186.3}{57.3}\right)}=17.6\times10^{-6} \tag{26}$$

The substitution of $\omega t_0=186.3°$ into equations 15 and 16 yields

$$\gamma=-E\frac{\omega t_0(K-1)\sin\omega t_0}{\omega t_0(K-1)(\sin\omega t_0+1)-\cos\omega t_0}$$

$$\gamma=-E\frac{\left(\dfrac{186.3}{57.3}\right)(3.5-1)\sin(186.3°)}{\left(\dfrac{186.3}{57.3}\right)(3.5-1)\sin(186.3°+1)-\cos(186.3)}=0.8E \tag{27}$$

and

$$I_b=\frac{\gamma C_1}{\omega t_0(K-1)}=0.099E\sqrt{\frac{C_1}{L_{ch}}} \tag{28}$$

γ is the initial voltage on capacitor C_1 at the beginning of the charging interval. Reference to Fig. 5 indicates that the maximum voltage appearing across the controlled rectifier is

$$V_{CR\ \max}=E+\gamma \tag{29}$$

Thus equation 29 determines the required voltage rating of the controlled rectifier. Using E equals 80 volts,

$$V_{CR\ \max}=E+\gamma=80+64=144\text{ volts} \tag{30}$$

Thus a controlled rectifier with a forward voltage rating of 150 volts was chosen for use in the experimental circuit.

Substitution of equations 27 and 28 in equation 8 results in an expression for current in the charging circuit.

$$i_{ch}(t)=1.8E\sqrt{\frac{C_1}{L_{ch}}}\sin\omega t+0.099E\sqrt{\frac{C_1}{L_{ch}}}(1-\cos\omega t) \tag{31}$$

Differentiating equation 31 and setting the resulting equation equal to zero yields

$$1.8\cos\omega t+0.099\sin\omega t=0 \tag{32}$$

Solution of equation 32 gives the time at which the current is a maximum. A value of ωt which satisfies equation 32 is

$$\omega t=93.2° \tag{33}$$

Substitution of equation 33 in equation 31 yields the following expression for the value of the charging current:

$$I_{\max}=1.9E\sqrt{\frac{C_1}{L_{ch}}} \tag{34}$$

or, solving equation 34 for $\sqrt{C_1/L_{ch}}$ gives

$$\sqrt{\frac{C_1}{L_{ch}}}=\frac{I_{\max}}{1.9E}=\frac{30}{(1.9)(80)}=0.1975 \tag{35}$$

The charging inductance L_{ch} and capacitance C_1 can be determined now from equations 26 and 35.

$$L_{ch}=89\ \mu\text{henrys} \tag{36}$$

$$C_1=3.48\ \mu\text{farads} \tag{37}$$

The average value of the charging current is given by

$$I_{ave}=\frac{1}{T}\int_0^{t_o} i_{ch}(t)dt \qquad (38)$$

Substituting equation 7 into equation 38 and performing the indicated integration yields

$$I_{ave}=\frac{1}{K\omega t_0}\left\{(E+\gamma)\sqrt{\frac{C_1}{L_{ch}}}(1-\cos\omega t_0)+ I_b(\omega t_0-\sin\omega t_0)\right\} \qquad (39)$$

Substitution of the values determined above yields

$$I_{ave}=5.45 \text{ amperes} \qquad (40)$$

Having completed the design of the charging circuit, it is now possible to proceed to the design of the magnetic discharge network. A 2-stage discharge network was chosen for the experimental pulse generator due to the simplicity of the resulting circuit. The volume per stage is given by:[1]

$$V_{ps}=\frac{\pi^2}{16}\,\frac{\mu_s}{B_s^2}\,U\left[\frac{2}{0.735\pi^2}\left(\frac{t_0}{\tau}\right)^2\right]^{1/} \qquad (41)$$

Orthonol was chosen for the magnetic material because of its square-loop properties and high saturation flux density. The saturated flux density of Orthonol is

$$B_s=1.41\times10^4 \text{ gauss} \qquad (42)$$

The saturated permeability is assumed to be

$$\mu_s=2.5\mu_0=31.7\times10^{-7}\ hy/m \qquad (43)$$

where μ_o is the permeability of free space. The peak energy per pulse stored in capacitor C_1 is $1/2C_1V_{C1p}^2$. The waveform of Fig. 5 indicates that this energy is

$$U=\frac{1}{2}C_1(2E+\gamma)^2 \text{ which equals}$$

$$\frac{1}{2}(3.48\times10^{-6})(224)^2=0.087 \text{ joule} \qquad (44)$$

Substitution of equations 42, 43, and 44 in equation 41 gives the volume per stage.

$$V_{ps}=7.96 \text{ cm}^3 \qquad (45)$$

The cores chosen for the magnetic discharge network were No. 51008-1/2*A* Orthonol cores made by Magnetics Inc. One-half-mil tape thickness was selected to keep the eddy currents small. The core dimensions are

$$V=8.1 \text{ cm}^3 \qquad (46)$$

$$A=0.65 \text{ cm}^2 \qquad (47)$$

$$\ell=12.49 \text{ cm} \qquad (48)$$

The volt-time integral appearing across saturable inductor P_1 is essentially $\int_{t_a}^{t_o} e_{c_1}(t)dt$. Use of equation 5 for $e_{C_1}(t)$ yields

$$\lambda_1=\int_{t_a}^{t_o}\left[E(1-\cos\omega t)-\gamma\cos\omega t- I_b\sqrt{\frac{L_{ch}}{C_1}}\sin\omega t\right]dt$$

or

$$\lambda_1=E(t_o-t_a)-\frac{(E+\gamma)}{\omega}[\sin\omega t_o-\sin\omega t_a]- \frac{I_b}{\omega}\sqrt{\frac{L_{ch}}{C_1}}[\cos\omega t_a-\cos\omega t_o]$$

t_a is the time at which the capacitor voltage passes through 0. Solution of equation 5 for ωt_a yields ωt_a equals 59.5°. Substitution of this value and the other known parameter values yields λ_1 equals approximately 6.0×10^{-3} volt-seconds.

The required number of turns for saturable inductor P_1 can now be calculated

$$N_1=\frac{\lambda_1}{2B_sA_1}=\frac{6.00\times10^{-3}}{2.82\times0.65\times10^{-4}}=33 \quad \text{turns} \qquad (49)$$

The first step in the design of the second stage is to calculate the saturated inductance of saturable inductor P_1 by use of equation 50.

$$L_{sn}=\mu_sN_n^2\cdot\frac{A_n}{\ell_n} \qquad (50)$$

$$L_{s1}=\mu_sN_1^2\cdot\frac{A_1}{\ell_1} =\frac{31.7\times10^{-7}\times33^2\times0.65\times10^{-4}}{12.49\times10^{-2}} =1.78\ \mu\text{henrys} \qquad (51)$$

If capacitance C_2, or its value reflected to the primary of saturable transformer P_2, is made equal to capacitor C_1, then the time required to transfer energy from C_1 to C_2 is one half the period of the resonant discharge of C_1 through C_2 and L_{s1} or

$$T_{12}=\pi\sqrt{L_{s1}\frac{C_1}{2}} =\pi\sqrt{1.78\times10^{-6}\times\frac{3.48}{2}\times10^{-6}} =5.55\ \mu\text{sec} \qquad (52)$$

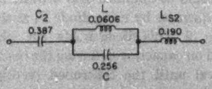


Fig. 7. Two-section Guillemin-type pulse-forming network. The element values are in henrys and farads and are normalized with respect to a load resistor of one ohm and a pulse width of one second

The integral of the voltage across the primary of saturable inductor P_2 is given by

$$\lambda_2=\frac{E_1}{2}\ T_{12}=\frac{224}{2}\times5.55\times10^{-6}=6.22\times 10^{-4} \text{ volt-sec} \qquad (53)$$

where E_1 is the peak value of the voltage across capacitor C_1 ($2E+\gamma$ or 224 volts).

The number of turns for the primary of saturable inductor P_2 is

$$N_{p2}=\frac{\lambda_2}{2B_sA_2}=\frac{6.22\times10^{-4}}{2.82\times0.65\times10^{-4}}=4 \text{ turns} \qquad (54)$$

Since two stages have been chosen for the magnetic discharge network, saturable transformer P_2 also serves as the output transformer to step up the voltage. A step-up ratio of 10 was chosen in order to to have a reasonably high output voltage and yet avoid high-voltage problems. Thus the secondary turns of saturable inductor P_2 and capacitance C_2 are given by

$$N_{s2}=40 \text{ turns} \qquad (55)$$

$$C_2=0.0348\ \mu\text{farads} \qquad (56)$$

To design the pulse-forming network, it is necessary to determine the saturated inductance of the secondary of saturable transformer P_2. Using equation 50, the saturated inductance is

$$L_{s2}=\mu_sN_{s2}^2\times \frac{A_2}{\ell_2}=\frac{31.7\times10^{-7}\times40^2\times0.65\times10^{-4}}{12.49\times10^{-2}} =2.64\ \mu\text{henrys} \qquad (57)$$

With this value known, the pulse-forming network may be designed. Fig. 7 shows a 2-section Guillemin-type network[7] used to discharge the last capacitor C_2 of the discharge network into the load. This network produces an approximate square wave of discharge current. The values shown are normalized to a load resistor of 1 ohm and a pulse width of 1 second. Because C_2 and L_{s2} are known, *L* and *C* may be determined by using the proportional values.

$$L=\frac{0.0606}{0.190}\times2.64\times10^{-6}\text{henry}=0.834\ \mu\text{henry} \qquad (58)$$

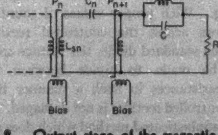


Fig. 8. Output stage of the magnetic pulse generator

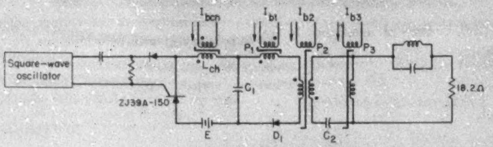


Fig. 9. Experimental pulse generator

$E = 80$-volt d-c supply
$L_{ch} = 89$ μhenrys
$C_1 = 3.5$ μfarads
$C_2 = 0.062$ μfarad
$D_1 =$ Transitron TR402 silicon rectifier
ZJ39A-150 = General Electric 150-volt controlled rectifier

$$C = \frac{0.256}{0.387} \times 0.0348 \times 10^{-6} \text{ farad}$$
$$= 0.0236\mu \text{ farad} \quad (59)$$

Fig. 8 shows the output stage of the generator. Nonlinear inductor P_{n+1} provides a low impedance path around R_L during the charging of C_n so that little power is lost in R_L during the charge. During discharge it assumes a large value of impedance.

Experimental Circuit

When the experimental pulse generator was constructed in the laboratory, various problems were encountered which are not considered in the analysis and design sections. These problems and their solutions are discussed in this section and the actual experimental circuit is presented.

One of the problems is associated with the "turn off" of the controlled rectifier. As previously mentioned, "turn off" time for the controlled rectifier is approximately 3 μsec. Even though this is a relatively fast "turn-off" time, the current in the charging circuit does reverse and reach an amplitude of about 4 amps in the negative direction before the controlled rectifier shuts off; see Fig. 10. The current drops abruptly to zero from this amplitude resulting in a large di/dt which, because of the charging inductor, produces a large spike of voltage across the controlled rectifier. This voltage can damage the controlled rectifier if it exceeds the peak inverse voltage rating of the device.

One solution to the problem is to add a conventional semiconductor diode in series with the controlled rectifier to obtain an increased PIV (variable speed) rating. With proper values of shunting resistors across the controlled rectifier and the standard diode, the voltage spike can be made to divide between the two resistances in such a manner that the controlled rectifier is not damaged. A more sophisticated method which does not require an additional rectifier is to wind the charging inductor on a magnetic core having square-loop characteristics. This core can be biased so that a positive current drives the core into its saturated region where it acts as a linear inductor, and a negative current drives the core into its unsaturated region. Since the core acts as a high impedance in its unsaturated region, it prevents the charging current from reversing significantly. Thus, little current exists when the controlled rectifier turns off, and the voltage spike across the controlled rectifier is reduced. Fig. 11 shows a photograph of the charging current when this type of charging inductor is used.

Another problem is the resetting of saturable inductor P_1. Inductor P_1 is reset by a constant bias current that produces a linear voltage build-up on capacitor C_1 in the negative direction. This voltage reaches a value γ when the core reaches negative saturation at the end of the resetting interval. Theoretically, the controlled rectifier is turned on at this time, and the charge stored on capacitor C_1 discharges through the controlled rectifier and the charging inductor and back on to C_1 in the positive direction.

However, if the circuit is operated slightly below the repetition rate it is designed for, the controlled rectifier is still off when saturable inductor P_1 reaches negative saturation and capacitor C_1 discharges through the magnetic discharge network in the negative direction. This pulse is clipped by the blocking reactor in the pulse-forming network so that some of the energy in the pulse is dissipated in the circuit. It was found that circuit operation could be obtained over a wider range of frequencies by using diode D_1 as shown in Fig. 9. This diode holds the charge stored on capacitor C_1 during the resetting interval until the controlled rectifier is turned on.

The experimental circuit with the charging inductor wound on a square loop magnetic core and a diode in the first loop of the magnetic discharge network is shown in Fig. 9. The triggering source for the controlled rectifier is a variable-frequency square-wave generator. The output is differentiated and the negative pulses clipped with a diode in order to obtain the positive pulses necessary to turn on the controlled rectifier.

Results

The results obtained from the experimental circuit are presented in this section. Waveforms of the charging current and voltage across capacitor C_1 are shown in Figs. 10, 11, and 12, respectively. These waveforms verify those predicted and have values approximately equal to the design values. The energy stored on capacitor C_1 can be calculated from Fig. 12 and is

$$U_a = \frac{1}{2} C_1 e_{C_1}{}^2_{\max} = \frac{1}{2}(3.5 \times 10^{-6})(210)^2$$
$$= 0.081 \text{ joule} \quad (60)$$

The predicted value given by equation 44 is

$$U_p = 0.087 \text{ joule}$$

so that the actual energy stored on capacitor C_1 is 0.006 joule less than the predicted value. At a frequency of 5 kc, this loss of energy corresponds to a loss in average power of 30 watts. The average power dissipated in the controlled rectifier, determined from General Electric Company data, is

$$P_{CR} = 7.9 \text{ watts} \quad (61)$$

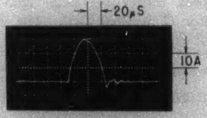


Fig. 10. Charging current when extra diode is used in series with controlled rectifier

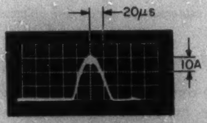


Fig. 11. Charging current using charging inductance with square-loop characteristics

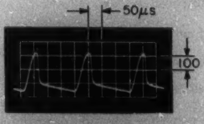


Fig. 12. Voltage across capacitor C_1

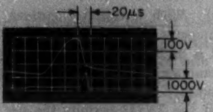


Fig. 13. Charge and discharge sequence of capacitors C_1 and C_2

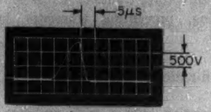


Fig. 14. Voltage across capacitor C_2

The copper loss in the charging circuit is given by:[2]

$$P_C=\frac{1}{2}(I_{\max})^2R'T_pf_r \qquad (62)$$

where

$I_{\max}$ = maximum value of the half sine wave current in amperes
R' = effective resistance in ohms
Tp = duration of the half sine wave current pulse in seconds
fr = repetition rate in cps (cycles per second)

Using the values given in Fig. 10 in equation 62, the average power loss in the charging circuit due to copper loss is

$$P_{Cm}=12.8 \text{ watts} \qquad (63)$$

The remaining loss of average power in the charging circuit is caused by core losses in inductor P_1 and capacitor losses. The core loss in inductor P_1 is

$$P_{p1}(\text{core loss})=10.1 \text{ watts} \qquad (64)$$

Fig. 13 shows the charge and discharge sequence of capacitors C_1 and C_2, and Fig. 14 shows the voltage across capacitor C_2. The energy stored on capacitor C_2 is determined from Fig. 14.

$$U_{C_2}=\frac{1}{2}\,C_2e_{C_2}{}^2{}_{\max}=\frac{1}{2}(0.062\times10^{-6})\times$$
$$(1.4\times10^{-6})(1.4\times10^3)^2=0.0606 \text{ joule} \qquad (65)$$

If the copper loss in the pulse-forming network is neglected, then the average power output at the load is

$$\overline{P}_o=f_rU_{C_2}=(5\times10^3)(0.0606)=303 \text{ watts} \qquad (66)$$

The loss in average power in transferring energy from capacitor C_1 to the load can be attributed to core, copper, and capacitor losses in the magnetic discharge network and pulse-forming network. Core losses can be estimated from experimental data take at the Brooklyn Polytechnic Institute.[3] Using the given data, the loss in average power due to the core losses of inductor P_2 and the blocking reactor P_3 are

$$P_{p2}(\text{core loss})=16.1 \text{ watts} \qquad (67)$$

$$P_{p3}(\text{core loss})=4.9 \text{ watts} \qquad (68)$$

From equation 62 the loss in average power due to copper loss in inductors P_1 and P_2 is

$$P_{p1}(\text{copper loss})=23.4 \text{ watts} \qquad (69)$$

$$P_{p2}(\text{copper loss})=9.3 \text{ watts} \qquad (70)$$

The average power dissipated in diode D_1 in the first loop of the magnetic discharge network is determined from manufacturer's data as

$$P_{D_1}=4.9 \text{ watts} \qquad (71)$$

The remaining power loss is due to losses in the capacitors.

Because no effort was made in the experimental circuit to minimize the bias currents, the effect of the bias currents is ignored in the calculation of efficiency. The measured average current in the charging circuit is

$$I_{ave}=5.35 \text{ amperes} \qquad (72)$$

An 80-volt d-c supply voltage was used so that the average power input becomes

$$P_{in}=EI_{ave}=(80)(5.35)=436 \text{ watts} \qquad (73)$$

Thus the efficiency of the magnetic pulse generator, ignoring the effect of the bias current, is

$$\eta=\frac{P_o}{P_{in}}=\frac{303}{436}=69.5\% \qquad (74)$$

The purpose of the investigation reported here was to explore a method for developing a 250-watt semiconductor magnetic pulse generator. The output power of the experimental circuit was approximately 300 watts. The large output power resulted from the use of a controlled rectifier as the electronic switch in the charging circuit. The controlled rectifier appears to be well suited for use as the electronic switch in this particular application because of its high power-handling capabilities, its ease of control, and its fast switching speed. Controlled rectifiers with voltage ratings up to 400 volts, or 250 volts higher than that used in the experimental circuit, are readily available. Thus, it appears that the output power of the type of magnetic pulse generator demonstrated in this investigation can be extended to higher values.

References

1. A TRANSISTOR-MAGNETIC PULSE GENERATOR FOR RADAR-MODULATOR APPLICATIONS, A. Krinitz. *Report no. 7848-R-1*, Massachusetts Institute of Technology, Cambridge, Mass., Sept. 1958.

2. THE USE OF SATURABLE REACTORS AS DISCHARGE DEVICES FOR PULSE GENERATORS, W. S. Melville. *Journal*, Institution of Electrical Engineers, London, England, vol. 98, Feb. 1951, pp. 185–204.

3. MAGNETIC MODULATORS FOR RADAR APPLICATION, E. J. Smith, J. Antin, K. T. Lian. *Report R-419-55, PIB-351*, Polytechnic Institute of Brooklyn, New York, N. Y., Apr. 1955.

4. P-N-P-N TRANSISTOR SWITCHES, J. L. Moll, M. Tanenbaum, J. M. Goldey, N. Holonyak. *Proceedings*, Institute of Radio Engineers, New York, N. Y., Sept. 1956, pp. 1174–82.

5. AVALANCHE BREAKDOWN IN GERMANIUM, S. L. Miller. *Physical Review*, New York, N. Y., vol. 99, Aug. 15, 1955, pp. 1234–41.

6. THE ELECTRICAL CHARACTERISTICS OF SILICON P-N-P-N TRIODES, I. M. Mackintosh. *Proceedings*, Institute of Radio Engineers, vol. 46, June 1958, pp. 1229–35.

7. PULSE GENERATORS (book), G. N. Glasco, J. V. Labacqz. McGraw-Hill Book Company, Inc., New York, N. Y., vol. V, 1948.

A Transistor-Oscillator Limit Switch for Indicating and Recording Instruments

J. T. WINTERMUTE
NONMEMBER AIEE

S. G. HAYTER
ASSOCIATE MEMBER AIEE

THE continued development of automatic controls has emphasized the need for limit switches on indicating and recording instruments which will not restrict the normal operation of the measuring unit. This control has been obtained in the past in two ways:

1. By contacts carried on the instrument pointer.
2. By employing a vane attached to the pointer to change the constants of a vacuum-tube oscillator circuit, and using the change in oscillator current to operate a relay.

Arrangement no. 1 has the disadvantage of being troubled with contact resistance unless magnetic contacts or a holding coil are used. With contacts carried on the pointer, the indication of the instrument is limited to the contact point setting. This is undesirable on recording-type instruments, where the purpose is to record the true value even when it is beyond the contact point setting.

The vane and oscillator arrangement avoids these shortcomings but introduces the element of limited life of vacuum-tube equipment. The arrangement to be described employs the vane and oscillator scheme but utilizes transistors, with their inherently longer life.

The transistor oscillator is changed from the oscillating to the nonoscillating state by the change in mutual inductance between the base and collector coils of an inductive feedback-type oscillator. This change in mutual inductance occurs when the moving element rotates a metal vane between the two closely coupled oscillator coils. When the vane is between the oscillator coils mutual inductance is decreased, oscillation stops, and the collector current drops to a low value, causing a relay to release.

A single-limit switch employs one oscillator and one set of feedback coils. A double-limit switch (a high and low limit) is achieved by employing two oscillator units.

A single-pole double-throw output relay is incorporated to provide either "make" or "break" contacts, depending on the terminals used.

Some of the major factors that had to considered in the development were sensitivity, i.e., the ability to secure control from a small motion of the moving element, self-starting of oscillation over a wide temperature range, and long life.

Oscillator Circuit Design

The oscillator employs a *2N332* silicon *n-p-n* transistor, which is a relatively low cost transistor with high storage and operating temperature limits. The oscillator circuit is a common-emitter, inductively coupled feedback type, with the base and collector coils mounted in the instrument. The *2N332* has a collector breakdown voltage of 45 volts minimum. It was desirable to keep the magnetic field set up between the collector and base coils as weak as possible, since strong fields would tend to oppose the vane motion as the vane approaches the "pull-out" position (the position where oscillation stops). With a collector current of 3 milliamperes or less, it was found that this effect is negligible even for small spring constants such as those used in small panel instruments. The high breakdown voltage of the *2N332* permits the use of a medium-sensitivity relay without resorting to large collector currents, by employing a high-resistance winding in the relay. A reverse-connected diode eliminates the induced voltage transient across the relay winding, preventing transistor breakdown.

Stabilization

Operation over a wide ambient temperature range of −20 to +50 degrees centigrade is achieved with essentially constant collector current by employing a thermistor to vary the forward bias applied to the transistor. The bias network shown in Fig. 1 consists of resistor R_1 and the shunt combination of resistor R_2 and thermistor *T*. R_1 is connected to the rectifier side of the filter resistor R_3. This results in a component being present in the bias voltage. The resulting collector voltage excursions assure self-starting at low temperatures. The operating point is selected to give class *B* operation, and with the "stiff" bias scheme shown the collector current change from the oscillating to the nonoscillating condition is very fast.

The power supply employs a constant-voltage transformer which permits reliable operation with line-voltage values from 80 to 150 volts, and provides isolation from the power line. Adjustable a-c degeneration, employing a variable iron-core inductance in the emitter lead,

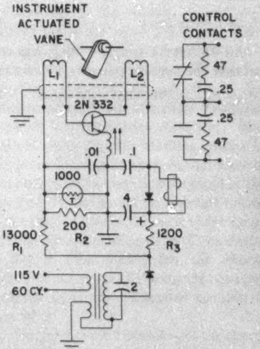


Fig. 1. Circuit diagram of transistor-oscillator control unit

MOVING VANE
COIL BLOCK
"LOW" LIMIT ARM
PEN POINTER
"HIGH" LIMIT ARM

Fig. 2. Application of transistor-oscillator high- and low-limit switches on a strip-chart recorder

Vol.
Pt.
196

Paper 59-160, recommended by the AIEE Recording and Controlling Instrumentation Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 1–6, 1959. Manuscript submitted October 24, 1958; made available for printing April 26, 1960.

J. T. WINTERMUTE and S. G. HAYTER are with the Westinghouse Electric Corporation, Newark, N. J.

The authors wish to acknowledge the early work of W. J. Schmidt, Westinghouse Meter Division, in applying transistors to the basic scheme of vane-controlled oscillators.

is used to adjust the over-all feedback. This adjustment is required to permit the use of commercial transistors with their rather large current gain spread. This also provides stabilization against change in transistor parameters with time. Shunt feedback between the collector and base could be used, but the tuned inductance method is more compact and readily locked in position, as compared with the variable resistor and blocking capacitor that would be required for shunt feedback.

Output Circuit

The present model employs a single-pole double-throw relay in the output with a contact rating of 5 amperes, 115 volts a-c noninductive load. A transistor switch or saturable reactor could be used instead, but would not give the independence of the controlled circuit that the relay provides.

Application

The transistor-oscillator limit switch was first applied to strip-chart recording instruments. The installation of oscillator coils on suitable adjustable supporting arms and a metallic vane on the moving element (Fig. 2) is feasible without upsetting the normal recorder functions. The recorder will operate over its full-scale range regardless of the limit device setting.

The small oscillator coils are rectangular in shape and are made with several hundred turns of very fine adhesive-coated wire. This type of wire is ideal for self-supporting coils and is readily cured by chemical softening of the wire coating and a subsequent short drying period. These coils are mounted in pairs in nonmetallic blocks with an open space between them. The metallic vane is so mounted on the moving element that the vane may pass freely through the gap between a pair of oscillator coils. The pen and vane move in the up-scale direction when the recorder is energized. Upon reaching a preset limit setting, the vane enters the gap between the oscillator coils and reduces the mutual inductance between the coils. The design of the vane assures continued interruption of the mutual inductance as long as the recorded value is in excess of the limit setting.

Rectangular-shaped oscillator coils are used in preference to round ones because rectangular oscillator coils will produce a greater mutual inductance change for a given angular rotation of the interrupting vane.

A step in the vane contour plus an exacting design of the oscillator coil support block and arms permits full-scale range for the limit switch settings. Recorders with two limit switches may set the switches with less than 5% of full scale between the high and low setting.

A multilead cable with a locking connecting plug at the external unit end is used for the circuits between the oscillator coils and the external oscillator unit.

Conclusions

Contact operation has been obtained from the moving element of a recorder or instrument without interference with its normal operation.

Transistors and semiconductor diodes have been used to provide long life and small over-all size.

Wire Line and Carrier Communications on the Union Pacific Railroad

C. O. JETT
MEMBER AIEE

COMMUNICATIONS in the area presently served by the Union Pacific Railroad Company preceded the Railroad by several years. The telegraph line of the Overland Telegraph Company, starting at Omaha, Nebraska, and extending westward through Salt Lake City, Utah, to San Francisco, California, was constructed in the years 1860 and 1861; while the construction of the Railroad did not start until 1864.

Inasmuch as the original telegraph line was hastily constructed, later, when it became necessary to rebuild, the line was moved onto the right-of-way of the Union Pacific Railroad and became the first wire communication facilities of the Railroad Company. Over the years more and more wire was added to the line for train dispatching, operation, and administrative purposes. Prior to 1912 only Morse telegraphy was used; however, in 1912, the first telephones were installed on the Union Pacific. The wire line plant reached its peak in miles of pole line and wire mileage in the early 1920's.

In the late 1920's voice repeaters were installed on a number of circuits to permit direct telephone communication from headquarters at Omaha to points as far west as Salt Lake City. Various types of printing telegraph machines were installed on the Union Pacific Railroad as early as 1918, and 4-channel Morkrum multiplex teleprinters were used as late as 1946, at which time all the teleprinter facilities were replaced with modern teleprinter equipment.

With the advent of carrier techniques and improved business conditions in the fall of 1937, a Western Electric type *CN-4*, 3-channel carrier system was furnished and installed by the Western Electric Company between Salt Lake City and Los Angeles, California. From this time on, the growth of business and the requirements for more and better communications necessitated rapid expansion of all forms of communications on the Union Pacific Railroad. Many single-wire Morse telegraph circuits were paired for telephone operation, voice transposed, and additional telephone instruments installed. Between a number of the more important offices, teleprinter facilities were expanded.

In 1941, just prior to the start of World War II, a Western Electric type *C-5*, 3-channel carrier system was installed between Omaha and Salt Lake City with terminals back to back at Cheyenne, Wyoming, to provide much needed additional facilities.

In rapid succession a 3-channel carrier system (Western Electric *C-5*) was installed between Salt Lake City and Portland, Oregon, and a second 3-channel system (Western Electric *C-5*) was installed between Omaha and Salt Lake City, to provide direct circuits from Omaha to Los Angeles, Salt Lake City, and Portland.

Restrictions on electronic equipment precluded any expansion of communications during World War II, except those that were mandatory for the handling of military operations. However, with cessation of hostilities, it became readily apparent that a major expansion must be made to provide administrative teleprinter facilities to consolidate the far-reaching points of the Union Pacific Railroad.

The new facilities for teleprinter service were provided by voice frequency carrier telegraph techniques. At the same time, many additional single-channel short-haul telephone carrier channels were provided on the Railroad by installing second and third story carrier systems over Western Electric's type *H-1* carrier systems which had been installed during the period 1939–1945.

Teleprinter Facilities

As mentioned previously, the Union Pacific Railroad first entered the field of voice frequency carrier telegraph facilities in 1946. In 1948, it was realized that many more channels for teleprinter service would be required than were available utilizing the conventional 170-cycle spacing for voice frequency carrier telegraph service. It was, therefore, decided to utilize equipment with 120-cycle spacing which would permit one-third more channels to be installed on a voice band. Further, a number of channels had been tried out in 1946 on what is now called "speech-plus-duplex." This technique involves removing that portion of the voice band between 1,450 and 1,900 cycles by suitable filters and installing two carrier telegraph channels within that space. The 120-cycle spaced channels then permit four carrier telegraph channels to be installed in the same space previously used by two 170-cycle carrier channels, thus doubling our teleprinter channel facilities. Today, a major portion of the Railroad's teleprinter facilities are derived by the "speech-plus-duplex" technique. There is only one location (Omaha–Salt Lake City) in which an entire voice band is utilized for voice frequency carrier telegraph service.

Several years ago, to provide additional facilities for teleprinter service, it was found that the only space available for expansion was in that band of frequencies lying above the voice range and below the "C" carrier systems, which has a frequency range of approximately 2,900–5,100 cycles. Twelve specially designed voice frequency carrier telegraph channels were then inserted into that space and today the Railroad operates 72 channels at various points in this band.

The present extent of the teleprinter service on the Union Pacific Railroad is shown by Fig. 1. It may be noted that the number of channels between various locations are not given; however, this factor is rather indefinite as channels are set up on patch basis from time to time depending upon service requirements. For locations not physically on the Union Pacific Railroad, circuits are obtained by lease from common carrier communication companies. There are a number of relay points to relay messages from one location to another. The torn-tape method is utilized to provide direct relaying without repunching a message. A typical teleprinter torn-tape relay center with six gang transmitter-distributors on the left and each cabinet on the right containing three typing reperforators is illustrated by Fig. 2.

Primary relay points are located at Omaha, Salt Lake City, Los Angeles, Portland, Chicago, and New York City. Secondary relay points are located at North Platte, Kansas City, Cheyenne, Pocatello, and Hinkle.

Typical teleprinter ending facilities are illustrated in Fig. 3.

Paper 60-815, recommended by the AIEE Wire Communication Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting, Atlantic City, N. J., June 19–24, 1960. Manuscript submitted December 21, 1959; made available for printing June 3, 1960.

C. O. Jett is with the Union Pacific Railroad, Omaha, Neb.

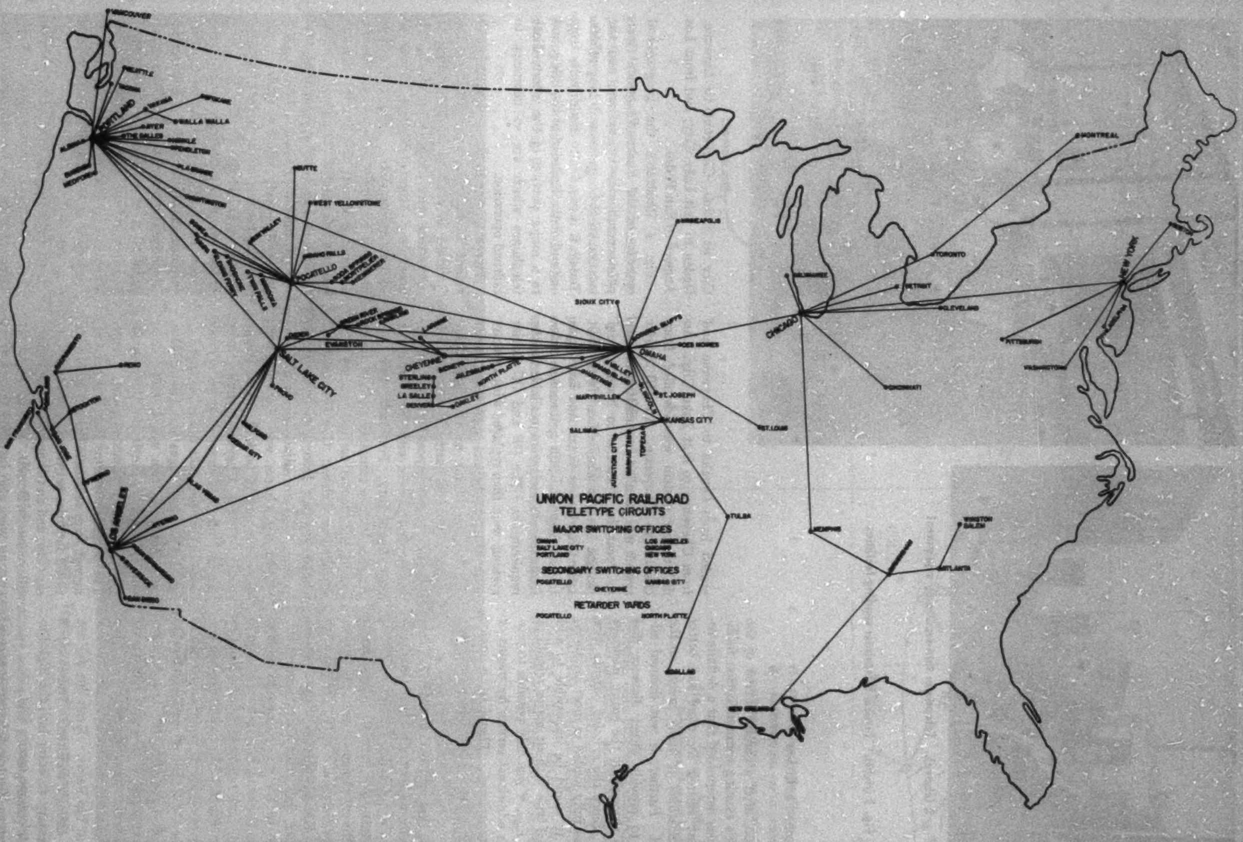

UNION PACIFIC RAILROAD
TELETYPE CIRCUITS
MAJOR SWITCHING OFFICES
OMAHA
SALT LAKE CITY
PORTLAND
LOS ANGELES
CHICAGO
NEW YORK
SECONDARY SWITCHING OFFICES
POCATELLO
CHEYENNE
KANSAS CITY
RETARDER YARDS
POCATELLO
NORTH PLATTE
VANCOUVER
SEATTLE
PORTLAND
SPOKANE
WALLA WALLA
THE DALLES
PENDLETON
LA GRANDE
BUTTE
WEST YELLOWSTONE
IDAHO FALLS
POCATELLO
MONTPELIER
TWIN FALLS
OGDEN
EVANSTON
GREEN RIVER
SALT LAKE CITY
PROVO
RENO
LAS VEGAS
LOS ANGELES
SAN DIEGO
CHEYENNE
STERLING
GREELEY
LA SALLE
DENVER
OAKLEY
JULESBURG
NORTH PLATTE
SIOUX CITY
COUNCIL BLUFFS
OMAHA
DES MOINES
VALLEY
GRAND ISLAND
LINCOLN
ST. JOSEPH
MARYSVILLE
KANSAS CITY
SALINA
JUNCTION CITY
MANHATTAN
TOPEKA
TULSA
DALLAS
NEW ORLEANS
MEMPHIS
ATLANTA
WINSTON SALEM
ST. LOUIS
MINNEAPOLIS
CHICAGO
MILWAUKEE
DETROIT
TORONTO
CLEVELAND
CINCINNATI
PITTSBURGH
WASHINGTON
MONTREAL
NEW YORK

Fig. 1. Union Pacific Railroad teletypewriter circuits

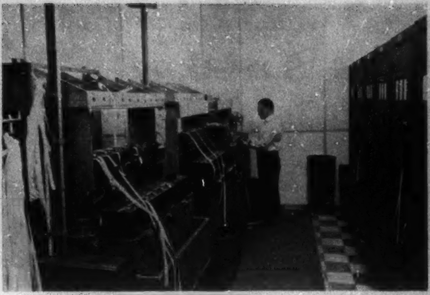


Fig. 2 (above). Teleprinter torn-tape relay equipment

Fig. 3 (right). Typical teleprinter sending facilities

Telephone Facilities

In 1955, after a careful review of the Railroad's existing communication facilities, it was determined that a definite requirement existed for more and better communications at many points on the Railroad. Further, it was deemed advisable to provide direct distance dial facilities throughout the entire Railroad system. Since the existing outside wire plant was fully loaded in most instances, in that it would be impossible to add additional wire without major rebuild of the pole line, it was only possible to provide this additional service by means of carrier techniques. Therefore, a major expansion into carrier facilities was started.

During the years 1955 to 1958, "C" carrier systems were installed from Omaha to Kansas City, Omaha to North Platte, North Platte to Cheyenne, Cheyenne to Green River, Green River to Ogden, Ogden to Salt Lake City, Cheyenne to Denver, Kansas City to Salina, Salina to Denver, Salt Lake City to Pocatello, Pocatello to Nampa, Nampa to Hinkle, Hinkle to Spokane, two systems Hinkle to Portland, and two systems Portland to Seattle. Further, because of the age and difficulty of obtaining maintenance material for the *C-4*, 3-channel carrier system between Salt Lake City and Los Angeles, it was replaced with a *C-5*, 3-channel carrier system extending from Salt Lake City to Los Angeles with terminals back to back at Las Vegas. Even this did not provide enough facilities. Therefore, it was necessary to enter into the high-frequency telephone carrier field.

A Western Electric *OB* system was installed from Omaha to Cheyenne and from Cheyenne to Salt Lake City. An *OC* carrier system was installed from Omaha to Cheyenne. An *OB* carrier system was installed from Salt Lake City to Pocatello, Pocatello to Nampa, and Nampa to Hinkle. Also, from Salt Lake City to Las Vegas, and Las Vegas to Los Angeles. Still not enough voice frequency channels were available for the requirements. The *OB* and *OC* carrier systems were then replaced with Lenkurt type *45-A* 12-channel open wire carrier installations. Prior to the installation of the 12-channel carrier systems, it was necessary to retranspose for 150-kc operation several pairs of wires from Omaha to Salt Lake City, Portland, and Los Angeles. Upon completion of the transposing, 12-channel carrier systems were installed as follows: One system direct Omaha to Salt Lake City with 16 repeaters. On the adjacent pair a second 12-channel system was installed with terminals back to back at Grand Island, North Platte, Cheyenne, and Green River. Also, 12-channel systems were installed between Salt Lake City and Las Vegas, and Las Vegas and Los Angeles, and another system was installed from Salt Lake City to Portland with terminals back to back at Pocatello and Nampa. The *O* carrier systems that had been removed were reinstalled to provide additional facilities between Omaha and Kansas City, and an *OB* and *OC* system from Cheyenne to Denver. Type *OB* carrier systems were also installed from Kansas City to Salina, Cheyenne to Laramie, Ogden to Salt Lake City, and from Los Angeles to San Pedro.

Fig. 4 illustrates the geographical coverage of the Railroad's present telephone communication system, and the location of the various switching offices. Today, it is possible to dial directly from one end of the Railroad to the other, and to a major portion of the intermediate telephones which are so necessary to railroad operation.

Presently, due to the shortage of competent Morse telegraph operators, the Railroad is engaged in a program of converting wire facilities on all branch lines from Morse telegraph service to telephone service. Many of these branches require ten or less telephones, permitting direct dial service through recently developed decimonic ringing techniques adapted for E and M lead control, while the remaining branches will utilize $3^{1}/_{2}$-cycle selector facilities for signaling because of the larger number of telephones involved.

The complexity and magnitude of the Union Pacific Railroad's communication system necessitate the use of commercial telephone company plant practices in the installation of equipment in the many locations where such facilities are required. Fig. 5, showing the communication office at Omaha, illustrates typical practices.

Due to the long association of the Union Pacific Railroad's communications department with the Western Union Tele-

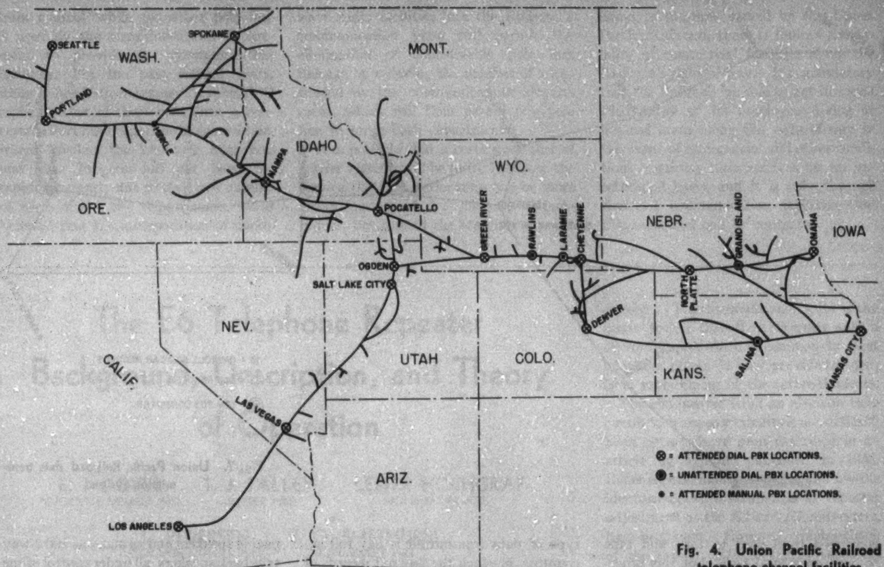

SEATTLE
WASH.
SPOKANE
MONT.
PORTLAND
HINKLE
NAMPA
IDAHO
WYO.
ORE.
POCATELLO
GREEN RIVER
RAWLINS
LARAMIE
CHEYENNE
NEBR.
GRAND ISLAND
OMAHA
IOWA
OGDEN
SALT LAKE CITY
NORTH PLATTE
DENVER
NEV.
UTAH
COLO.
CALIF.
KANS.
SALINA
KANSAS CITY
LAS VEGAS
ARIZ.
LOS ANGELES
⊗ = ATTENDED DIAL PBX LOCATIONS.
● = UNATTENDED DIAL PBX LOCATIONS.
• = ATTENDED MANUAL PBX LOCATIONS.

Fig. 4. Union Pacific Railroad telephone channel facilities

graph Company, many Western Union practices have been continued, especially with respect to testing and patching the open wire plant. A typical Railroad testing and patching switchboard installation is shown in Fig. 6.

Data Facilities

In 1956 the Railroad entered into a program to expedite movement of trains and the handling of identifying correspondence by employing techniques called electronic data processing methods, or, in simple language, "machine accounting." Equipment manufactured by IBM (International Business Machines Corporation) was utilized. Originally, data were transmitted on teleprinter channels and

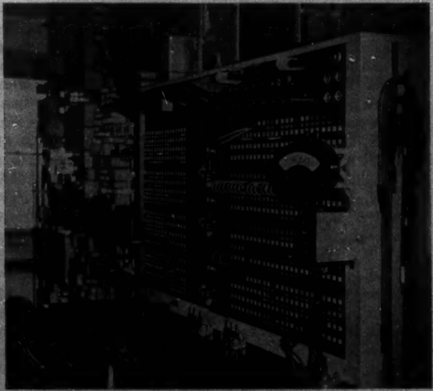


Fig. 5 (left). Typical communication equipment rack

Fig. 6 (above). Typical wire testing and patching facility on a railroad communication system

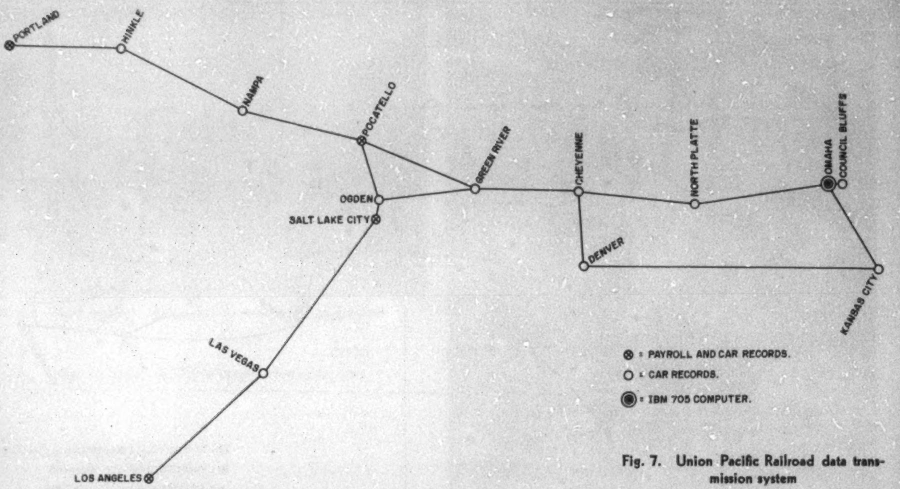


Fig. 7. Union Pacific Railroad data transmission system

converted from card to tape and tape to card on a local basis, but this procedure was cumbersome, wasteful of manpower, and much too slow.

The program of data transmission was speeded up in 1957 by the elimination of the intermediate process of card to tape and tape to card by utilizing the transceiver manufactured by IBM, which transmitted a card from one transceiver to a similar transceiver at the distant end on 100-wpm teleprinter channels. After approximately a year's experience with this type of equipment, it was realized that it, too, was inadequate for the assigned task, and that it would be necessary to use higher speeds which necessitate using the voice frequency type transceiver, on a 4-wire basis, whereby four transceiver channels can be derived from one voice frequency telephone channel, resulting in the doubling of the number of cards per minute transmitted and received by each transceiver circuit.

Early in 1958, a comprehensive data system was established over the entire Railroad to feed data from the various originating points to the IBM Model *705* computer located at Omaha, for car record operation between the various reporting offices. This system is shown in Fig. 7.

Today, two types of data are being handled on the present facilities. One type is the car record, which is the operational record of all freight cars moving throughout the Union Pacific system, and to connecting railroads. The second type of data transmitted is pay roll accounting, in which the pay roll records of the employees of the Railroad are channeled daily to the computer in Omaha, where the pay roll record is kept and the checks prepared for payment on a semimonthly basis.

Specialized Facilities

Only in railroad operation is it necessary for one person to direct and control an entire facility covering, in many instances, several hundred acres. This is the condition that exists in railroad yards. To expedite the movement of trains and the switching of cars through the various terminals, it has been necessary to establish large intercommunicating and paging facilities whereby the supervisor can contact and direct many different persons, diesel switching crews and repair forces. This is accomplished by means of strategically located paging speakers and 2-way talk-back speakers which permit an employee to signal his supervisor, or the supervisor to speak to personnel even when they are several hundred feet distant from the speaker. Each of these yard intercommunicating systems is custom engineered to provide the best coverage in view of the service requirements. This service in the yards is generally provided through buried multiconductor telephone cables comparable in many cases to the telephone exchange plant in villages and small towns. Yards frequently have 15 to 25 paging speakers and as many as 100 2-way talk-back speakers, all under control of one supervisor. A similar problem exists in the modern terminal freight house with incoming and outbound freight and transshipment of freight from one car to another, in that the entire operation must be under the direct control of one individual. The problem most generally is solved with a system somewhat smaller than that utilized for yard operations.

Future

Presently, there is no further room for expansion on the wire line communications plant and yet the requirements exist for more facilities for voice communications, teleprinter channels, and for the transmission of data of many types and from many places on the Union Pacific into Omaha. The cost to expand the wire communications plant would be prohibitive. The Railroad, therefore, is entering into a program of installing a microwave radio communication system. Presently under construction is a system from Omaha to Laramie, and the section from Laramie to Salt Lake City is now in the planning stage. It is hoped that the communications facilities so derived by microwave techniques will be adequate for a number of years.

Presently over the major portion of the Railroad two pole lines exist. One pole line is used by the signal department to control the operation of signals and in certain areas to provide centralized train control, while the other pole line is used by the communications department to provide the communication facilities. For the past several years, when it has been necessary to rebuild or replace either of these pole lines, a concerted effort has been made to consolidate, where feasible, the two pole lines into one line. Progress has not been too rapid, frequently due to the large number of wires of the two departments. It is believed that the incorporation of microwave radio facilities into the Railroad's communication plant will permit the elimination of considerable pole line mileage by reducing the number of wires needed by the communications department, which will then permit one pole line to serve both departments.

It is possible that a certain number of carrier systems will be retired because the service they are performing can be more easily performed by microwave techniques, but due to the scarcity of population in the area served by the Union Pacific Railroad, there is limited availability of commercial facilities along the Railroad's right-of-way. It is mandatory that the Railroad be able to get in touch with many of its employees living at isolated areas along the right-of-way in the event of emergencies and other conditions requiring their services on an unscheduled basis, and it is only through Railroad communication facilities that this personnel can be contacted.

The E6 Telephone Repeater Background, Description, and Theory of Operation

E. J. DOYLE
ASSOCIATE MEMBER AIEE

T. J. TALLEY
MEMBER AIEE

LESTER HOCHGRAF
NONMEMBER AIEE

L. PEDERSEN
ASSOCIATE MEMBER AIEE

J. O. SMETHURST
NONMEMBER AIEE

IN 1948 the Western Electric Company introduced a 2-way telephone voice repeater which operated in a radically different manner than had previous 2-way telephone repeaters. The new repeater employed the negative impedance principle, meaning that it was essentially a 2-terminal device intended to be inserted in series with a telephone circuit. Previous voice repeaters were 4-terminal units made up of combinations of 1-way amplifiers and hybrid coil circuits connected to provide 2-way transmission. The popularity of the new repeater stemmed from its low cost and from the fact that it could pass d-c supervisory currents and dial pulse signals without need for special bypass equipment.

The first of the series, the *E1*,[1] and the later *E2* were inserted in series with the line pair and they introduced a negative resistance which increased the voice currents at both the receiving and sending ends of the connection. The former action was desirable since it increased the received signal, but the latter caused echo if the repeater was used on connections with appreciable transmission time delay. The large impedance discontinuity introduced by the repeater limited the operating net loss of the circuit to avoid singing when the circuit was in the idle condition and caused poor frequency response when the circuit was in use.

Another negative impedance device, the *E3* repeater, was developed to be used specifically in combination with the series repeater. It was connected across the cable pair, not in series with it. The combination series-shunt *E13* or *E23* repeater could be adjusted to provide a much better impedance match to the line pair, thus reducing reflection and echoes and permitting operation of the circuits at lower net loss.

The manufacture and use of *E*-type repeaters was confined to a few thousand a year for several years after their introduction in 1948. However, with the need for improvement in transmission of circuits in metropolitan areas, the number installed mushroomed rapidly. By early 1959 more than one million units of the *E1*, *E2*, and *E3* had been built and installed in the Bell System.

The maintenance experience on these repeaters over a 10-year period has been very favorable. Few of the repeaters have failed, and routine maintenance is limited to a semiannual or annual check of the electron tube cathode activity. Such checks are easily and quickly made. The stability of gain and gain-frequency characteristics has been good. What, then, are the characteristics that would be desired in a new repeater?

Transistorization of the *E2* and *E3* is one obvious step. The *E23* repeater uses two dual electron tubes, requiring a total of 7 watts of 48- and 130-volt supply. Transistorization cuts the power to less than 2 watts using only a 48-volt supply, which is common to most telephone offices. It may provide freedom from replacement of the active elements.

The changeover from an electron tube circuit to transistor circuits is not difficult. Such circuits have been described in an article by Linvill[2] published in 1953. Experimental models, which could use the identical network strappings or repeater adjustment as the *E2* and *E3* converters, have been built. However, one drawback of the present *E13* and *E23* repeaters is the method of adjustment of the repeater, both to match line impedance and to provide the desired gain over an adequate frequency band. Once the repeater has been adjusted to work with a given line it has proved to maintain its adjustment very well, but the initial procedure is time consuming and it must be repeated if the repeater gain is to be changed because of reassignment of the cable facilities. A new repeater should therefore minimize the adjustment procedure.

Operation and Gain Adjustment of E23 Repeater

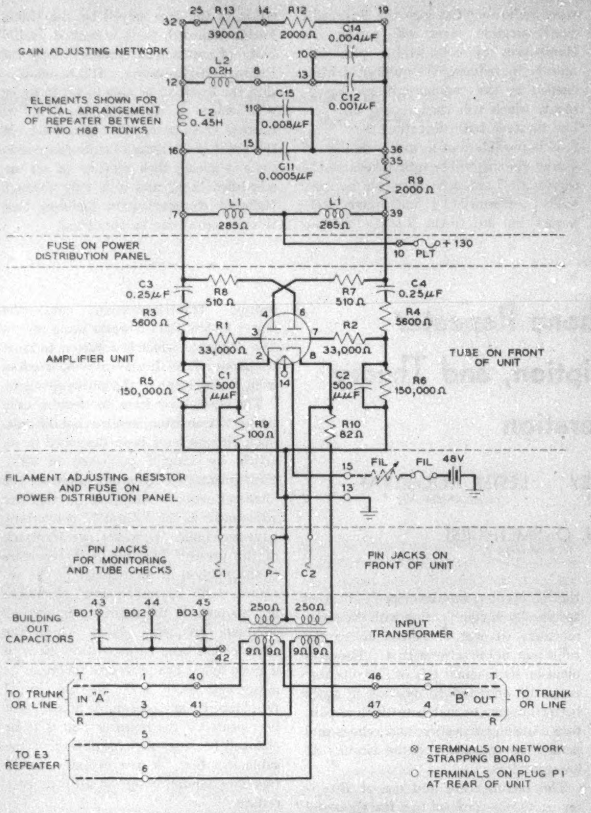
A brief description of the method of operation and adjustment of the *E23* re-

---

Paper 60-814, recommended by the AIEE Wire Communication Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting, Atlantic City, N. J., June 19–24, 1960. Manuscript submitted March 24, 1960; made available for printing May 4, 1960.

E. J. Doyle and T. J. Talley are with the American Telephone and Telegraph Company, New York, N. Y.; Lester Hochgraf is with the Bell Telephone Laboratories, Murray Hill, N. J.; and L. Pedersen and J. O. Smethurst are with the Bell Telephone Laboratories, North Andover, Mass.

Important contributions to the development of the line-building-out network were made by Miss R. L. Huxtable, American Telephone and Telegraph Company, and by J. L. Garrison and R. W. DeMonte of the Bell Telephone Laboratories; to the development of the converters and test sets by A. L. Bonner and W. J. Kopp of the Bell Telephone Laboratories. Figs. 31 and 32 are reproduced by courtesy of J. W. Topping of the New Jersey Bell Telephone Company.



GAIN ADJUSTING NETWORK
ELEMENTS SHOWN FOR TYPICAL ARRANGEMENT OF REPEATER BETWEEN TWO H88 TRUNKS
R13 3900Ω
R12 2000Ω
C14 0.004µF
L2 0.2H
C15 0.008µF
C12 0.001µF
L2 0.45H
C11 0.0005µF
R9 2000Ω
L1 285Ω
L1 285Ω
FUSE ON POWER DISTRIBUTION PANEL
10 PLT
+130
C3 0.25µF
C4 0.25µF
R8 510Ω
R7 510Ω
R3 5600Ω
R4 5600Ω
R1 33,000Ω
R2 33,000Ω
AMPLIFIER UNIT
TUBE ON FRONT OF UNIT
R5 150,000Ω
C1 500 µµF
C2 500 µµF
R6 150,000Ω
R9 100Ω
R10 82Ω
FILAMENT ADJUSTING RESISTOR AND FUSE ON POWER DISTRIBUTION PANEL
FIL
FIL
48V
PIN JACKS FOR MONITORING AND TUBE CHECKS
C1
P-
C2
PIN JACKS ON FRONT OF UNIT
BUILDING OUT CAPACITORS
BO1
BO2
BO3
250Ω
250Ω
INPUT TRANSFORMER
9Ω 9Ω 9Ω 9Ω
TO TRUNK OR LINE
IN "A"
"B" OUT
TO TRUNK OR LINE
TO E3 REPEATER
T
R
⊗ TERMINALS ON NETWORK STRAPPING BOARD
○ TERMINALS ON PLUG P1 AT REAR OF UNIT

Fig. 1. Schematic diagram of E2 repeater

peater will serve to point out the functions of various parts of the negative impedance circuit and the repeater-adjustment problem. This brief exposition will preface the description of the operation of the *E6* repeater.

A schematic diagram of the *E2* series repeater is reproduced in Fig. 1 from the article by Merrill, Rose, and Smethurst.[3] Note the passive elements which are shown on the top of the diagram above the dashed line. This is the variable network or adjustable portion of the series repeater. Impedance of this network is transformed to a negative value by the tube converter in the amplifier unit. The line terminals, which serve as both input and output terminals, are connected to the cable pair through the input transformer shown at the bottom of the diagram. In most applications the repeater is used on coil-loaded cable pairs between two central offices. The gain-adjusting network shown in Fig. 1 is for one particular value of gain with the repeater connected to two *H-88* loaded cable pairs which have 0.5 length end sections; i.e., the nearest loading coils are 3,000 feet away from the repeater.

The negative impedance and gain inserted in the line are not varied by changing the conversion ratio of the amplifier portion, but by changing the magnitude of the positive impedance of the variable network. The line impedance to which the repeater should be adjusted is shown in Fig. 2(A) for both 19- and 22-gage *H-88* loaded cable pairs with 0.5 length end sections at the repeater.

End section lengths different from the above are taken care of by varying some of the capacitance components in the network, or by building out short-length end sections (less than 3,000 feet) with the capacitors labeled *B01*, *B02*, or *B03*, shown near the lower part of Fig. 1. This complexity of adjustment is necessary because the impedance of loaded cable pairs is a complex function of frequency over the voice band; it also varies with the length of end section as shown in Fig. 2(B). Similar adjustments are required in the *E3* repeater network.

Each different gain setting of the repeater requires changing the values of most, if not of all, components in both *E2* and *E3* networks. The desired configuration and component values are obtained by strapping wire connections and soldering them to terminals on the sides of the *E2* and *E3* units. The 39 network-adjusting terminals of an *E2* repeater connect to 12 fixed resistors, 11 capacitors, and 3 tapped inductors. Similarly, in an *E3* unit there is a nearly equal number of network components.

A test set has been devised to speed the strapping adjustment prior to permanently wiring and soldering a repeater network, but the procedure is time consuming and inflexible. The strapping procedure requires much time during intial installation or during the subsequent eassignment of an *E2* or *E23* repeater.

Considerable engineering time has also been spent on working out the best network strappings for a given repeater gain and cable loading. Because of this work, and because of the field work required to change to a different gain value, these data have been issued only in 0.5-db (decibel) gain steps for series repeaters used alone, and in approximately 0.8-db steps for *E23* repeaters.

Method of Operation of E6 Repeater

It has been recognized for some time that the method employed in the *E2* and *E3* repeaters, of providing a large number of network elements and using relatively few components for any particular installation, is wasteful of apparatus and engineering, and of installation time. Evidence from the field indicates that only a few typical network arrangements of the many configurations possible are used in a majority of cases. To confirm these conclusions, a survey was made to determine the types of exchange-area facilities in common usage. From analysis of the data it was discovered that about 85% of all trunks were *H-88* loaded cable circuits in either 19-, 22-, or 24-gage cable.

Vol.
Pt. 1
1960

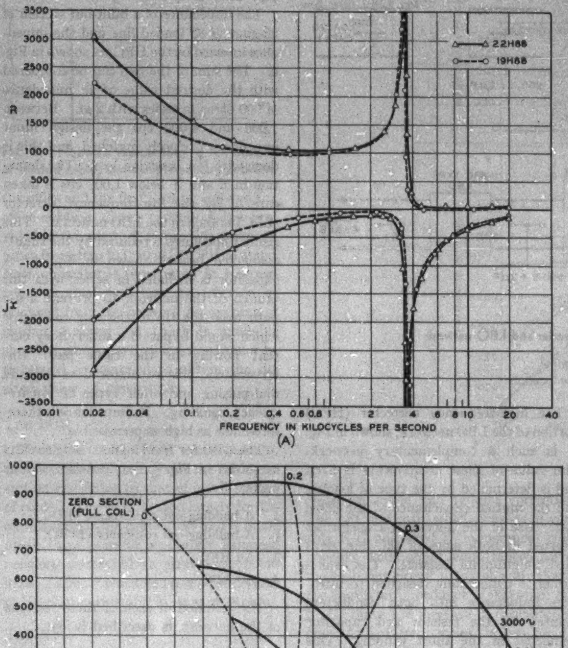


Fig. 2. Characteristic impedances of H-88 cable pairs

A—Of 19 and 22 gage terminated at mid-section
B—Of 24 gage for various end sections

Furthermore, *H-88* loading is the only widely used arrangement recommended for future additions to the exchange-area plant. As a result of this investigation it was decided to develop a new repeater for *H-88* loaded cable circuits and to let future demands determine the necessity for other types of networks.

The field of use of the *E6* repeater is similar to that of the *E23* in short-haul trunks and special-service lines wherever moderate gains suffice. Its freedom from 130-volt power needs, however, suits it ideally to many situations where provision of power for *E23* repeaters would have been too costly. It will be widely used in trunks between offices in the same exchange area: local-to-local, local-to-tandem, and local-to-toll.

In the *E6* repeater the functions of impedance matching and gain production have been separated. Instead of matching the repeater to the line, the line is matched to the repeater. This may seem trivial but it results in a repeater which is easier to adjust and which has better transmission performance. The impedance transformation is made by a passive network called the line-building-out network (LBO), which is connected between the loaded line and the gain unit as shown in Figs. 3(A) and (B). This network transforms the impedance presented by the loaded pair for different end section lengths and wire gages to a constant value which, in the case of *H-88* loading, is made equivalent to an impedance of 900 ohms in series with 2 μf (microfarads). This impedance at the line terminals having been obtained, the complex gain network of the *E2* and *E3* repeaters can be replaced by a much simpler structure. The adjustable gain network is reduced to an adjustable resistor which is relatively easy to change.

Figs. 3(A) and (B) show the schematic connection of an *E6* repeater to lines at a circuit terminal and also at an intermediate office. The LBO network is always connected between the gain unit and cable pairs. One LBO is required for a terminal repeater and two for an intermediate repeater. On the office side of a terminal repeater a dummy LBO furnishes the connections between the converter and the line contacts on the plug.

LBO Network

The input impedance of an *H-88* loaded cable pair was shown as a function of frequency in Fig. 2(A) and as a function of both frequency and end section length in Fig. 2(B). A function of one part of the *E6* LBO network is to build

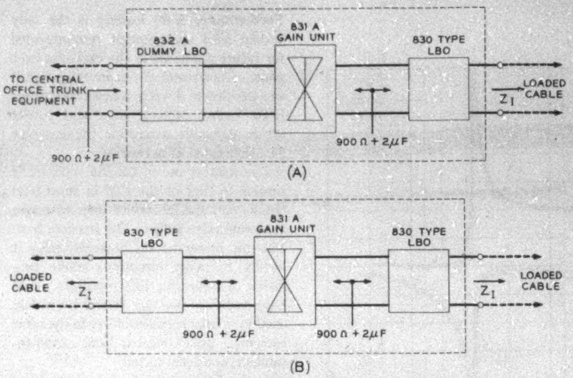

832 A
DUMMY LBO
831 A
GAIN UNIT
830 TYPE
LBO
TO CENTRAL
OFFICE TRUNK
EQUIPMENT
LOADED
CABLE
Z_I
900 Ω + 2μF
900 Ω + 2μF
(A)
830 TYPE
LBO
831 A
GAIN UNIT
830 TYPE
LBO
LOADED
CABLE
Z_I
Z_I
LOADED
CABLE
900 Ω + 2μF
900 Ω + 2μF
(B)

Fig. 3. Arrangement of E6 converter and LBO network

A—At circuit terminal
B—At intermediate location

out or extend the cable end section to give the same impedance as that of a full-length end section (6,000 feet for *H* loading). This is done with shunt capacitors and series resistors shown on the left side of Fig. 4.

The impedance of a full-length end section of loaded cable at frequencies above about 1,000 cps (cycles per second) resembles the impedance of a constant-*K* low-pass filter. This impedance can be simulated closely by the network shown on the left side of Fig. 5. It is possible to design a complementary network that can be inserted in series with the full-length end section impedance to achieve a nearly constant input resistance over a wide frequency band. This type of network is shown on the right side of Fig. 5.

The high-frequency corrector (HFC) portion of the LBO network, shown in Fig. 4, is such a complementary network. The value of these components is fixed and is determined by the type of loading and the mutual capacitance of the cable pair. Use of the 3-winding transformer inserted in both sides of the line gives good longitudinal balance. The transformer is deliberately designed to introduce inductance into both conductors shunted by the resistor and capacitor connected in the third winding. This part of the LBO network presents a nearly constant resistive impedance to the gain unit at frequencies above the cable cutoff frequency where the full-section cable pair impedance becomes very low. The reactance of this portion of the network corrects the reactance of the cable pair impedance at, and just below, the loaded pair cutoff frequency.

The impedance of a built-out section of 22-gage *H-88* loaded line and the impedance inserted by the HFC are shown in Fig. 6. The sum of the two can be compared with the desired value of an impedance of 900 ohms in series with 2 μf. Between 1,000 and 4,000 cps the design input impedance is closely matched, and this is done with low insertion loss. The deviation in *R* and *X* below 1,000 cps is taken care of by the low-frequency corrector (LFC) branch of the LBO network. This rising impedance is reduced by the shunt-connected branch shown in Fig. 4. A capacitor is included in series with this branch of the network to prevent a d-c path from the tip to the ring conductor which would bypass d-c supervisory current flowing in the cable pair. Unfortunately, this capacitor does degrade dial-pulsing and other types of low-frequency signaling. Its impedance is therefore made as high as permissible.

The LBO for *H-88* loaded cable circuits as shown in Fig. 4 thus consists of four parts:

1. A building-out capacitance (BOC).
2. A building-out resistance (BOR).
3. An HFC.
4. An LFC.

The function of each of these sections of the network is described below.

THE BOC

The BOC network is connected across the cable pair, at the repeatered point, for building out the cable end section capacitance to that of a full section of cable. It can compensate for irregularities between

Vol
Pt.
196

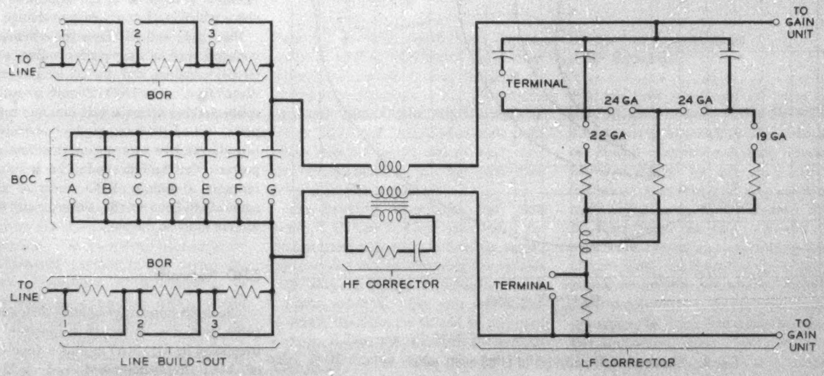

TO
LINE
1
2
3
BOR
BOC
A
B
C
D
E
F
G
BOR
TO
LINE
1
2
3
LINE BUILD-OUT
HF CORRECTOR
TO
GAIN
UNIT
TERMINAL
24 GA
24 GA
22 GA
19 GA
TERMINAL
TO
GAIN
UNIT
LF CORRECTOR

Fig. 4. Schematic of an LBO network

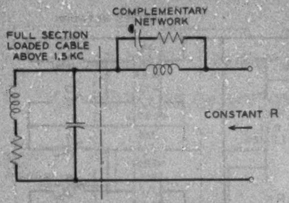


Fig. 5. Network representing a full section of loaded cable, and complementary networks

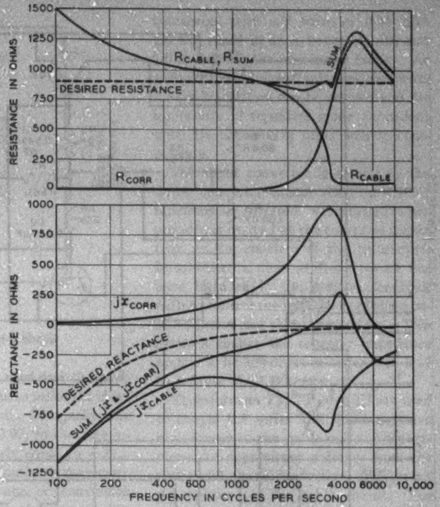


Fig. 6. Impedance characteristic of loaded cable pair and with LBO network

the office and the first loading coil and for office cabling. Seven capacitors are provided in the network to provide a capacitance from zero to 0.10 μf in 0.001-μf steps. Each capacitor is terminated in a screw-type terminal and may be inserted or removed from the network by opening or closing a screw. A chart is available showing which particular screws must be tightened or loosened to obtain the desired value of capacitance.

The BOR

The BOR resistors are connected into the network for building out the cable end section resistance to that of a full section of cable. The BOR is used largely for 22- and 24-gage cable conductors and is made up of three pairs of accurately balanced resistors connected to screw-type terminals. The six resistors have a total value of 196 ohms adjustable in steps of 28 ohms.

The HFC

The HFC network inserts series impedance to make the cable, as built out to a full section by the BOC and BOR, look like a constant resistance through and beyond cutoff. In balanced circuits two precise series networks would be required to preserve the line balance for longitudinal currents. Through the use of transformer techniques, a high degree of balance is secured by a single *R-C* (resistance-capacitance) network. Such an arrangement results in reduced size of components by appropriate application of the impedance transformation available in the design of a 3-winding transformer. Since the same value of series impedance is required for all gages of cable of a given capacitance and loading, no adjustment for gage is provided in the HFC network.

The LFC

The impedance characteristic resulting from the addition of BOC, BOR, and HFC is one that approaches a constant resistance above frequencies of 1,000 cps. However, at frequencies lower than 1,000 cps the resistance and capacitive-reactance terms of the cable impedance increase in value and require correction. The LFC network, which consists of variable R-C and fixed *L* (inductance) components, adds shunt admittances across the built-out circuit to match more closely the converter impedance of 900 ohms and 2 μf. Since the increase in cable impedance, at low frequencies, is less for 19-gage than for 22- and 24-gage cable, the 19-gage corrector has the highest impedance and the smallest effect. Three adjustments are provided in the LFC network to assure the optimum adjustments for 19-, 22-, and 24-gage pairs.

When the LBO is associated with intermediate repeaters, less than full LF correction is applied because of the adverse effects of two networks on both signaling

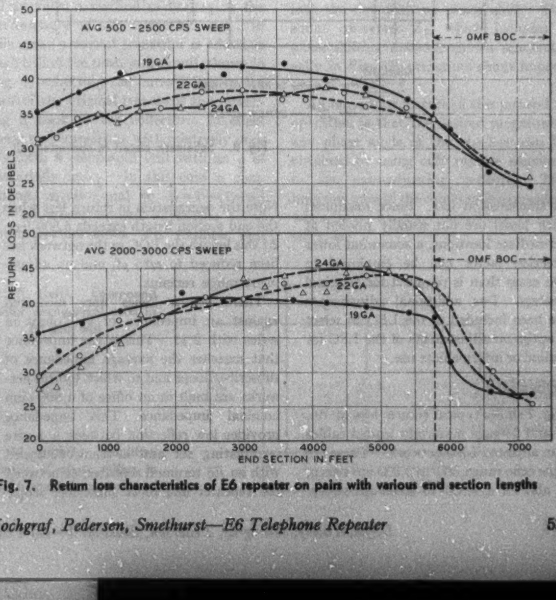


Fig. 7. Return loss characteristics of E6 repeater on pairs with various end section lengths

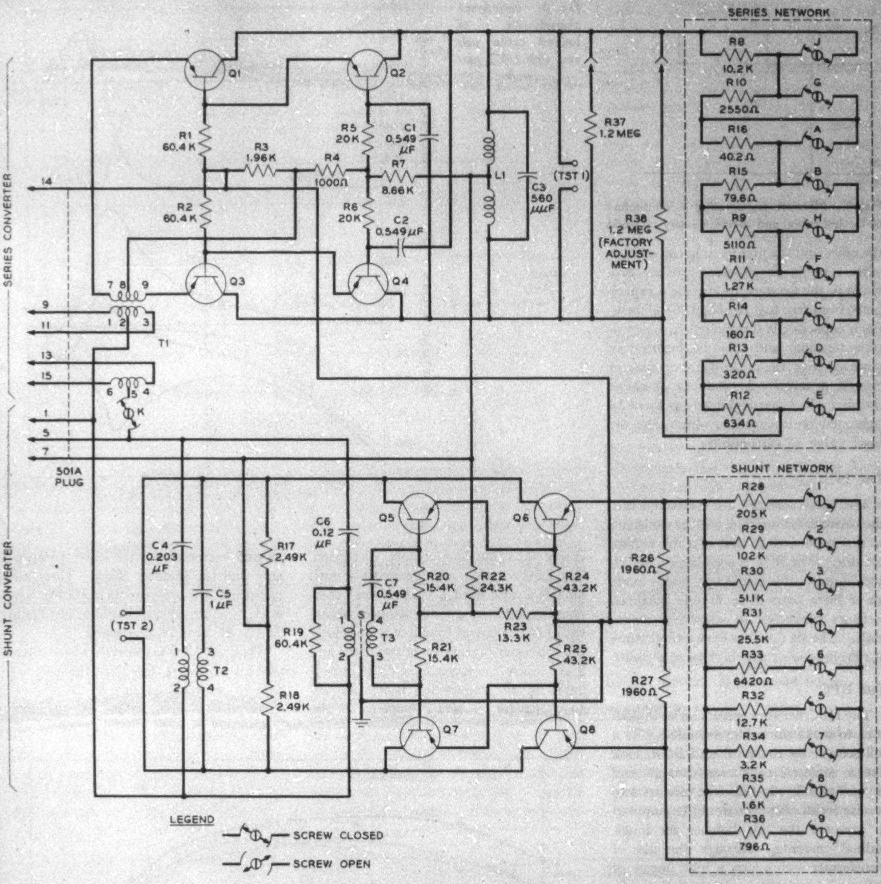

SERIES NETWORK
SERIES CONVERTER
SHUNT CONVERTER
SHUNT NETWORK
Q1
Q2
Q3
Q4
Q5
Q6
Q7
Q8
R1 60.4 K
R2 60.4 K
R3 1.96 K
R4 1000Ω
R5 20K
R6 20K
R7 8.66 K
C1 0.549 μF
C2 0.549 μF
L1
C3 560 μμF
(TST 1)
R37 1.2 MEG
R38 1.2 MEG (FACTORY ADJUSTMENT)
R8 10.2 K
J
R10 2550Ω
G
R16 40.2Ω
A
R15 79.6Ω
B
R9 5110Ω
H
R11 1.27 K
F
R14 160Ω
C
R13 320Ω
D
R12 634Ω
E
14
7 8 9
1 2 3
T1
9
11
13
15
6 5 4
K
1
5
7
501A PLUG
R28 205 K
R29 102 K
R30 51.1 K
R31 25.5 K
R33 6420Ω
R32 12.7 K
R34 3.2 K
R35 1.6 K
R36 796Ω
1
2
3
4
6
5
7
8
9
R26 1960Ω
R27 1960Ω
C4 0.203 μF
C5 1 μF
(TST 2)
T2
1 2 3 4
R17 2.49 K
R18 2.49 K
C6 0.12 μF
C7 0.549 μF
R19 60.4 K
T3
1 2 S 4 3
R20 15.4 K
R21 15.4 K
R22 24.3 K
R23 13.3 K
R24 43.2 K
R25 43.2 K
LEGEND
SCREW CLOSED
SCREW OPEN

Fig. 8. Schematic circuit of converter unit

and transmission loss. Since maximum return losses are not usually needed at intermediate locations, a somewhat lower correction factor can be employed in these cases than is required for terminal repeaters. Two additional adjustments have been included in the LBO to select the optimum adjustments of the LFC for terminal or intermediate use.

Performance

Fig. 7 shows typical return loss in 19-, 22-, and 24-gage uniformly loaded cables for an adjusted LBO network. These are for the echo range, 500- to 2,500-cps sweep, and also for a 2,000- to 3,000-cps sweep. Note the degradation in return loss when the end section length exceeds 5,800 feet. At this length the BOC in the network has been reduced to zero μf, and no adjustment range remains.

These return losses were measured against an impedance of 900 ohms in series with 2 μf. This is the impedance that matches the average impedance of subscriber loops and to which hybrid networks are built in an office of a 900-ohm nominal impedance. This impedance provides low reflection between a 4-wire terminating set and a trunk equipped with an *E6* terminal repeater, or between the repeater and most subscriber loops. Thus low reflection is obtained either with the repeater located at the point of connection to an intertoll trunk or with the repeater located at the end office.

Gain-Unit or Converter Circuits

The gain unit is the active or amplifier portion of the repeater. It corresponds to a 4-terminal gain pad adjustable to an image impedance of 900 ohms in series with 2 μf and from zero to 13 db in gain. As with the *E23* repeater, the gain unit consists of a series-connected and a shunt-connected converter in a bridged-T connection. A common coupling transformer

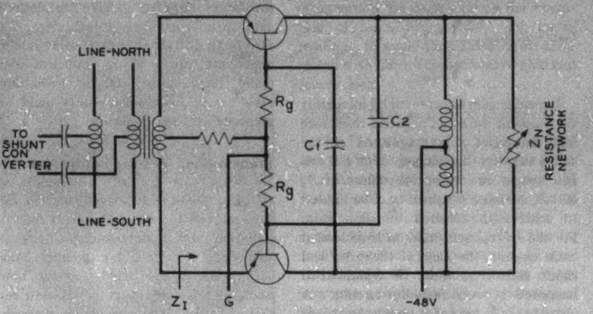


Fig. 9. Simplified schematic of series negative resistance converter circuit

having an impedance ratio of 1 to 9, is used for joining the series and shunt elements together to form a 4-terminal network. The negative series impedance is introduced into the line by a pair of balanced windings on the line transformer and the shunt admittance is bridged across the mid-points of the two line windings. A schematic circuit of the gain units of the *E6* repeater is shown in the diagram of Fig. 8.

The negative impedance series converter is a 4-terminal network consisting of a transistorized positive-feedback amplifier, designed to be open-circuit stable, and a variable resistance network.

A simplified schematic of the series negative converter is shown in Fig. 9. The series-connected converter is a 4-terminal circuit whose impedance at the "input" terminal *T1* is the negative of the value of adjustable resistance connected in the network. This type of circuit using transistors has been described by Linvill.[3]

When C_1 and C_2, the feedback capacitors, are made large and have low impedance relative to R_1, the ratio of E_1 to I_1, which is the input impedance, is given by

$$\frac{E_1}{I_1}=(1\text{-}2\alpha)Z_{\text{net}}+2r_{\text{emit}}+(1-\alpha)2r_{\text{base}}$$

The resistance of r_{emit} is a few ohms at the emitter current value used and the ratio of Z_{net} to r_{base} is such as to make the third term a few per cent. Ordinarily the transistor α is about 0.97, so that the input impedance is nearly equal to $-Z_{\text{net}}$. Since the α's of good transistors may vary, the negative impedance or conversion ratio is dependent upon the proper selection of transistors. In typical units, which have initial α's between 0.95 and 1.0, the 5% spread between units would result in a 12% change in the ratio of Z_{in} to Z_{net}. Possible aging of the transistors would further degrade the performance, and this extent of variation is intolerable when the design requirements allow a maximum deviation in gain of only ±0.2 db for each converter.

The compound transistor circuit suggested by Darlington[4] and shown in Fig. 10 provides a method of improving the performance of the converter unit. In this array a second transistor is added to each half of the push-pull circuit, shown in Fig. 9, collecting the current loss in the base of the original transistor and returning it back to the collector. In effect these two transistors with relatively poor α's form a compound unit with an α of essentially unity. In this case α compound, α_c, is equal to $\alpha_c=1-(1-\alpha_1)(1-\alpha_2)$, where α_1 and α_2 are the α's of the individual units. For example, new transistors with α's between 0.94 and 1.0 will result in an α_c between 0.996 and 1.0 to stabilize the conversion factor. As a result, the initial and aging requirements for the transistors can be relaxed and the uniformity of repeaters made dependent upon the precision of other passive components.

The shunt converter is similar to the series converter except that the circuit functions are reversed. The negative impedance is derived at the collectors rather than at the emitters of the transistors. In addition, a phase-shifter network has been included in the positive feedback path to control the impedance of the combined series and shunt converters. This network makes the repeater impedance appear to be 900 ohms in series with 2 μf to balance the LBO network.

Transformers *T1*, *T2*, and *T3* are used to couple the series and shunt negative impedances to the cable pair. They remove d-c supervisory voltages from the series and shunt converters, and permit d-c pulsing on the cable circuit. The transformer *T1* functions as a bridged-T structure with both the image impedance and gain controlled by varying the values of the series arm Z_A and the shunt arm Z_B.

If the image impedance of the repeater is made equal to the image impedance of the line, as modified by the LBO network, no reflections will occur. The circuit gives gain in both directions of transmission.

In the *E6* repeater the series converter Z_A inserts a negative resistance of $-R$ into the line with no phase shift. The shunt converter Z_B, which includes a phase-shifting network shown schematically in Fig. 11, presents a shunt impedance of $Z\underline{/180-\theta}$.

An important feature of this method of operation is that the desired magnitudes and phase angle of the repeater can be obtained by using only resistor elements in the gain-adjusting networks. The insertion gain of the combination repeater, when facing its image impedance on both sides, is

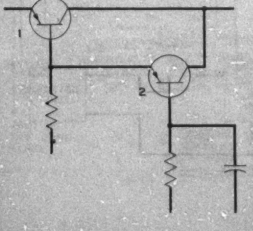


Fig. 10 (left). Compound transistor arrangement used in E6 gain unit

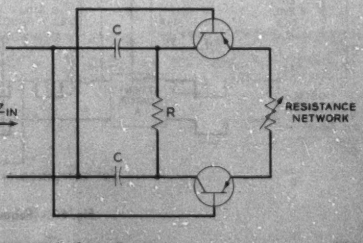


Fig. 11 (right). Simplified schematic of phase shifter in E6 shunt converter

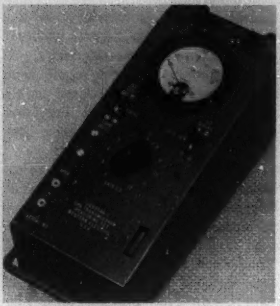


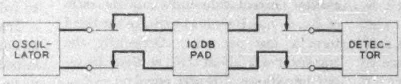
Fig. 12. The 54A transmission measuring set

$$\text{Gain} = 20\log_{10}\frac{1+\sqrt{\dfrac{Z_A}{4Z_B}}}{1-\sqrt{\dfrac{Z_A}{4Z_B}}}$$

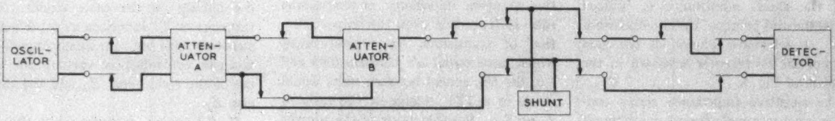
Repeater gain is increased by increasing the value of the series negative resistance and decreasing the magnitude of the shunt negative impedance. For a specified image impedance, the values of Z_A and Z_B are fixed and must be co-ordinated to achieve the desired insertion gain. For the *E6* repeater, used on *H-88* loaded cable circuits, the gains of the series and shunt converters must be equal when measured between 900-ohm circuits. A chart has been prepared giving the values of the series and shunt converter network resistances required to give the desired gain, in approximately 0.1-db steps, for zero to 13 db. If circuits of other image impedance are ever in sufficient demand, for example, 600- or 1,350-ohm cable circuits, additional charts could be prepared for these facilities.

Transmission Testing Equipment

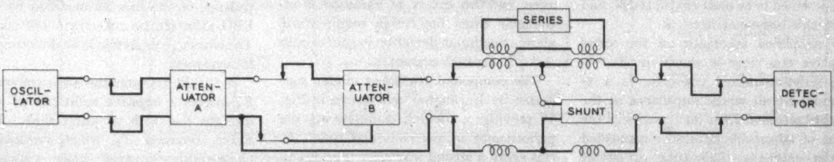
One of the aims for developing the *E6* repeater was to reduce the costs for installation and maintenance. Considerable simplification and economies have been realized by designing three new pieces of test equipment for the new repeaters. The *54A* transmission measuring set is used to determine the gain of


OSCILLATOR
10 DB PAD
DETECTOR

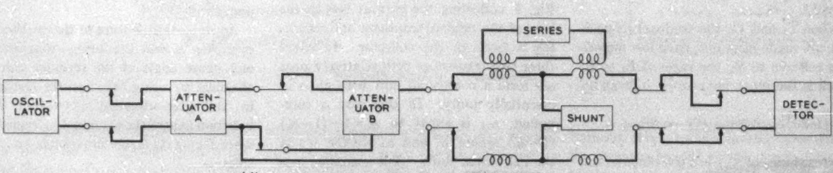
(a) CALIBRATE


OSCILLATOR
ATTENUATOR A
ATTENUATOR B
SHUNT
DETECTOR

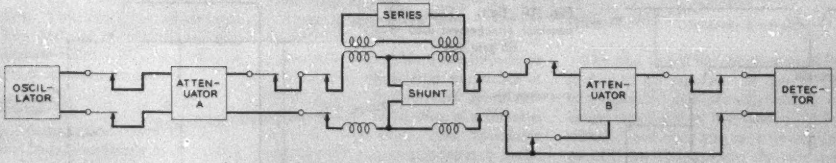
(b) SHUNT CONVERTER GAIN


SERIES
OSCILLATOR
ATTENUATOR A
ATTENUATOR B
SHUNT
DETECTOR

(c) SERIES CONVERTER GAIN


SERIES
OSCILLATOR
ATTENUATOR A
ATTENUATOR B
SHUNT
DETECTOR

(d) COMBINED SHUNT AND SERIES CONVERTER GAIN


SERIES
OSCILLATOR
ATTENUATOR A
SHUNT
ATTENUATOR B
DETECTOR

(e) COMBINED SHUNT AND SERIES CONVERTER LOAD TEST

Fig. 13. Sequence of transmission and load measurements

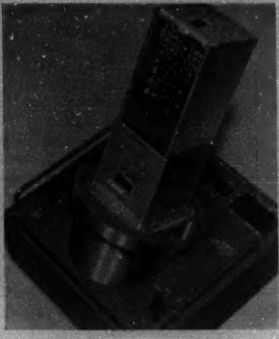


Fig. 14. The 54B test stand

the series and shunt converters, individually or combined, to insure that the proper impedance relationships have been set up to obtain high return losses between the converters and LBO networks. A companion unit, the *54B* test stand, provides a convenient device for holding the repeaters in an upright position for gain adjustments; it includes test jacks and keys for establishing connections to make return-loss measurements. The third unit, coded as the *54C* return-loss measuring set, is arranged to measure the return losses of the cable circuits, the cable circuit plus LBO, and the cable circuit plus LBO plus converters against a termination of 900 ohms and 2 μf, when optimum performance is required of the *E6* repeater. All three sets are designed for simplicity of operation, by co-ordinating the operation of keys, attenuators, and meter readings, to avoid the human errors associated with the measurements of gain and return loss.

THE 54A TRANSMISSION MEASURING SET

The *54A* set is used to adjust the gain and the image impedance of the converter or gain-unit section of the *E6* repeater. It is a portable transistorized device, shown in Fig. 12, which is used not only for initial repeater adjustments, but also for making periodic measurements of gain and signal-level carrying capacity of the converters to detect any deterioration of the transistors. All keys and switches are arranged so that measurements can be made quickly, to reduce repeater out-of-service time.

The transmission set consists of a 1,000-cps oscillator, two attenuators, a detector, and a meter circuit, designed to measure the gain of the series and shunt converter circuits between precise 900-ohm impedances. By measuring the insertion gain of series and shunt converters separately, and then adjusting these gains for equality, an image impedance of 900 ohms plus 2 μf is obtained for the combined unit.

The oscillator consists of an *R-C* type of bridge circuit, using two transistors, with the output level stabilized by including a small tungsten filament lamp in one arm of the bridge. A third power transistor, with a 12-db local negative feedback, acts as a buffer stage to furnish a +13 dbm output power for measuring purposes. The output impedance is adjusted, at the factory, to be within 900± 3 ohms.

The detector is a transformer-coupled, push-pull, class *B* transistorized amplifier with the indicating meter inserted into the common battery lead. The meter movement is reverse-biased, to provide a linear scale feature, and a readable scale range of 3 db is used for gain measurements. The meter face plate is calibrated to read zero at mid-scale for an input signal of +3 dbm, and covers a range of ±1.5 db in 0.1-db steps. Approximately 40 db of negative feedback is used to stabilize transistors and guarantee meter accuracy. Internal compensation is included in the set to correct for changes in transistor characteristics with temperature. Since the over-all accuracy must be better than 0.05 db to give reliable readings of 0.1 db, an internal calibration circuit is provided to take care of small deviations in the oscillator and detector output levels. The power supply, obtained from the central office 48-volt battery, is stabilized by means of a Zener diode regulator. The input impedance is adjusted to be 900±3 ohms at 1,000 cps.

A switching circuit and two built-in attenuators are used to measure gain and overload of the *E6* converters. Fig. 13 shows the connections used in calibrating the *54A* set and in measuring gain and overload with it. Repeater gains between zero and 13 db can be measured. For any gain the switching circuit will insert sufficient attenuation ahead of the converters to read an output of +3 db referred to one milliwatt. Upon operation of the load-measuring key, the input attenuator is

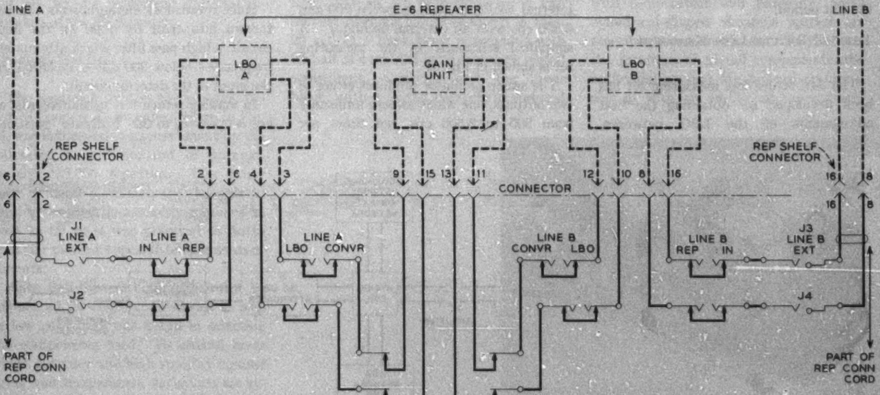


Fig. 15. Test connections provided by 54B stand

transferred to the output circuit for applying a high-level signal to the converters. With good transistors, the gain should not change more than the amount prescribed in the operating instructions. Larger reductions in gain indicate transistor impairment and probable cause for replacement of the devices.

The 54B Test Stand

The *E6* repeater cannot be adjusted on the relay-rack mounting shelf because, in this position, the adjusting screws for the LBO and converter networks are inaccessible. The *54B* test stand, shown in Fig. 14, has been developed to hold the repeater in place, to power it for installation and maintenance tests, and to provide convenient connections to various points for return-loss and trouble-location tests. The test stand consists of a rotating platform, turnable through 180 degrees; one position provides access to the converter unit adjusting screws, and the opposite side provides access to the LBO network adjusting screws. A test cord, plugged into the repeater-shelf connector jack, extends the line connections to the *54B* set. Forty-eight-volt power for operating the repeater is obtained from a socket provided on the relay-rack power distribution panel, and multiple outlets are available in the test stand for providing power to operate the *54A* and *54C* equipments. Fig. 15 shows the location in the circuit of the jacks provided for testing purposes.

A holding coil circuit is provided which gives a d-c path from conductor to conductor, on either side of the repeater. This prevents a previously dialed-up test termination from being released during the test period.

The 54C Return-Loss Measuring Set

The *54C* return-loss measuring set has been developed for obtaining the best adjustments of the LBO networks. Much of the improved performance of the *E6* repeater depends upon the optimum settings of these networks. Limited experience on several field trials indicates that the cable records do not always contain precise knowledge of the end section resistance and capacitance. The records appear to underestimate the capacitance of the end section by amounts which range from several hundred to 1,500 feet in length.

Adjustments made with the aid of the *54C* set compensate for the true end section resistance and capacitance and make it possible to secure better return loss for mixed gages and deviations in cable characteristics in the end sections. It is designed to give an integrated measure of return loss of the LBO network, repeater, and cable against a standard termination in the echo and singing bands. While its primary purpose is to determine the most favorable settings for the network adjustments, it can be employed to give a quantitative measure of swept band return loss. Ultimately, it may find broader applications in making similar adjustments of impedance compensators and precision networks associated with 2-wire toll-connecting trunks and for cable acceptance tests. Jacks are provided so that an external oscillator may be used with the hybrid coil and detector portions of the set.

The *54C* set, shown in Fig. 16, is a return-loss measuring device consisting of a sweep oscillator, a detector and a meter, a single-coil hybrid, and a compromise network of 900 ohms in series with 2.16 μf. Return losses from -13 to $+45$ db can be measured over the frequency band from 500 to 2,500 cps with the internal oscillator, and between 200 and 3,300 cps with an external oscillator. A simplified schematic of the measuring set is shown in Fig. 17.

The sweep oscillator produces either of two outputs, one which sweeps uniformly from 500 to 2,500 cps, ten times per second, and the other which sweeps from 2,000 to 3,000 cps at the same rate. The output signals are amplified by a 2-stage transistorized circuit to produce the desired sweep frequency test power.

The 2,000- to 3,000-cps sweep is used for adjusting the BOC of the LBO network which is most critical at the higher audio frequencies. The 500- to 2,500-cps sweep is used to adjust the BOR and LFC portions of the LBO network.

The hybrid coil is a 600- to 10-ohm ratio transformer which will hold calibration throughout the range of return-loss values measured. Terminals are available on all four sides of the hybrid coil for connection to external test equipment, such as a single-frequency oscillator, to detectors and to networks. The transformer is not designed to permit direct current to flow through the line windings. The internal 900-ohm plus 2-μf network has no d-c path and therefore can have a signal voltage present across the cable pair without having direct current flow through the hybrid coil.

The detector is a 4-stage transistorized amplifier and rectifier whose output is fed into a d-c meter. Approximately 45 db of negative feedback is used to stabilize the transistors and to improve the linearity of rectifier and meter. An input attenuator, calibrated in 10-db steps, is used to keep readings on the meter scale. The meter is shunted by a capacitor which integrates the received power throughout each 1/10-second sweep to produce a steady meter reading. Response of the detector is approximately linear to changes in the level of the input signal, and it measures the average return loss of the frequency band being swept.

Since return-loss measurements can be thrown into error by noise on the line circuit, a high-pass filter which attenuates frequencies below 200 cps is included in the input of the detector circuit.

In making return-loss measurements, a key is operated to the "calibrate" position

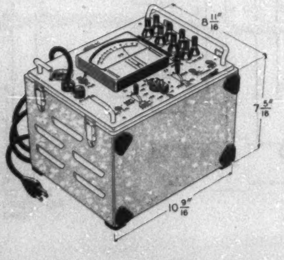


Fig. 16. The 54C return-loss measuring set

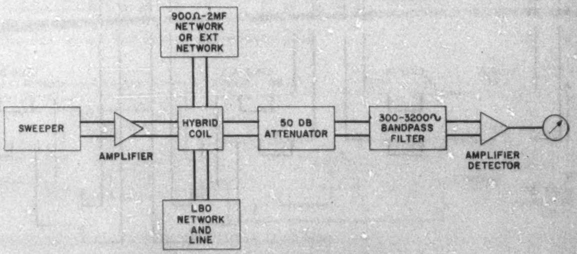


Fig. 17. Simplified schematic of 54C set

Vol.
Pt.
196

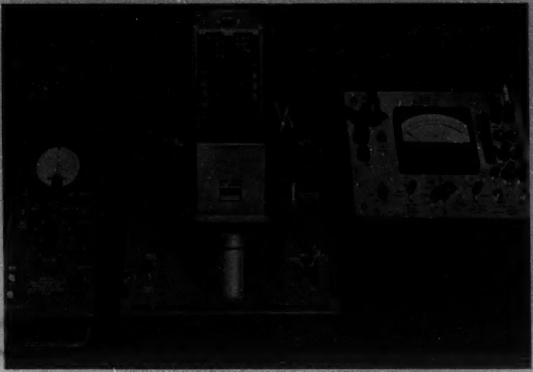


Fig. 18 (above). Typical test arrangement for E6 repeaters

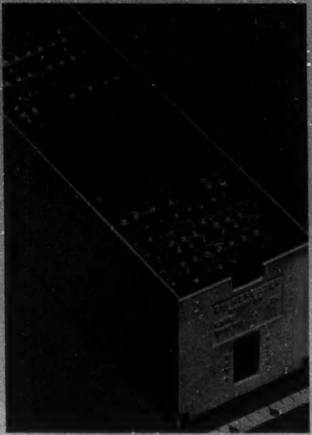


Fig. 19 (right). E6 repeater for intermediate use

and the output of the sweep oscillator is adjusted for a zero-db reading on the meter scale. Restoring the key to the "measure" position and adding or removing loss from the input attenuator to bring the meter reading on scale will give the return loss as the difference between two readings. Attenuator and meter readings in red indicate negative return losses; black readings are used for positive return losses.

The *54C* is powered from the 48-volt office battery obtained from the power-distribution panel or from a multipled outlet in the *54B* test stand. Fig. 18 shows a typical arrangement of the *54A*, *54B*, and *54C* sets for measuring gain and return loss of the *E6* repeaters.

Physical Design

The approach used in the physical design of the new *E6* repeater, shown in Fig. 19, was influenced to a great extent by the estimated annual demand of approximately 125,000 repeaters. This fairly large demand permitted the designer to consider automatic assembly methods at least as far as it was practical to do so with our present types of telephone components.

Since Bell System equipment must be engineered for continuous service as well as low cost, care was taken in selecting the components used. In critical areas where accuracy and high stability against aging and temperature variations are required, small wire-wound resistors with 1/2% and 1.0% tolerances are specified. In other cases, it was satisfactory to use deposited-carbon types, mostly with 1.0%, but some with as high as 5.0%, tolerances.

Metallized mylar capacitors were selected because of their accuracy, small size, and low cost. The line coil is a new design precision transformer with provisions for adjustment of both inductance and d-c resistance to close limits.

The transistors employed are of the Western Electric Company alloy p-n-p germanium type.

The method selected for mounting this variety of components is the so-called matrix structure or sandwich type. The structures of the converter and the LBO network are each made up of two molded phenolic blocks as shown in Figs. 20 and 21. Each block is provided with its half of a molded cavity for each pigtail component. Consider first the lower block: with a pigtail component inserted in its proper cavity, the one lead wire is extended through a hole in the bottom of the cavity and then through a printed wiring board which is attached to the bottom side of the block. When all pigtail components have been placed in this manner, the second or top block is fitted to complete the sandwich structure. The top block, molded with the proper cavities, is also equipped with a printed wiring board which will receive the second lead wire of the pigtail components. When the matching halves of the phenolic block are put together, they form a closed housing for the pigtail components.

To compensate for dimensional changes with temperature and humidity, small cylindrical-shaped neoprene spacers are placed between the blocks. To facilitate the insertion of pigtail components, the individual cavities are tapered or funnel shaped with an additional tapered sec-

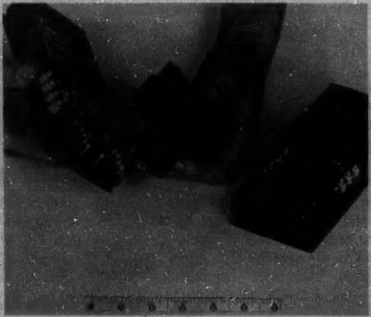


Fig. 20. Converter assembly

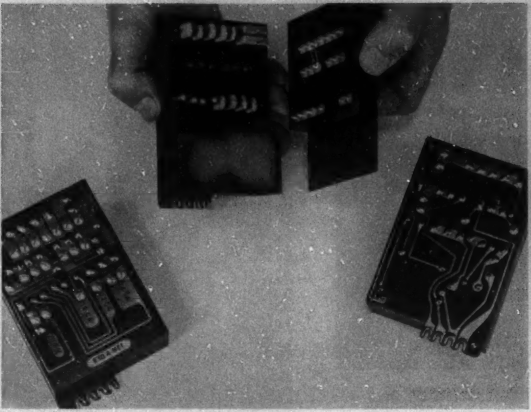


Fig. 21. LBO network

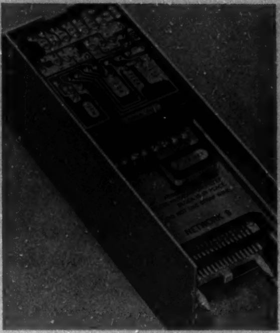


Fig. 23. The E6 repeater for terminal use

tion at the base to guide the pigtail lead wire through to the printed wiring board. The pigtail leads are precut to proper length and no clinching is performed for connections to the printed conductor. All connections are dip soldered.

Components, other than pigtail types, such as transformers and inductors, are assembled in open cavities that are entirely part of the bottom block without a matching top section. Following test and adjustment of these components, the cavities are filled with a microcrystalline wax and protected with a metal plate in the case of the converter and with a printed wiring board for the LBO network. The transistors are an exception to the cavity method of mounting because of the problem of heat transfer, and to facilitate replacement. The shelf arrangement on the lower block of the converter provides a mounting surface for the transistors, as shown in Fig. 20.

Each adjusting screw in the converter and LBO networks is provided with a captive washer which bridges or opens a gap between two printed wiring conductors. To insure reliable contacts the captive washers are gold plated and the printed conductors have a hot tinned finish. To guard against vibration, a nylon nut, located in the molded cavity of the top phenolic block, is used for the adjusting screw.

When the converter and the associated LBO networks have been assembled in their matrix structures, the electric connections between the two are made by means of screws. The comblike protrusion in the phenolic block on one end of the LBO network, shown in Fig. 21, consists of four open-end brass grommets, soldered to the printed wiring and arranged to receive the screws that go into tapped metal bushings located on the converter. The bushings have a terminal arrangement that permits soldering connections to the printed wiring.

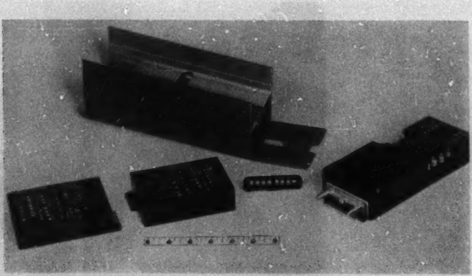


Fig. 22. Aluminum housing and networks

The printed wiring boards used in this assembly are made from a paper-base phenolic, grade *XXXP*, 1/32 inch thick and clad with 2-ounce copper on one side. The conductor width is 1/16 inch and a minimum spacing of 0.050 inch between adjacent conductors has been specified. To maintain a higher order of insulation resistance, two spray coats of clear lacquer are applied to the exposed board surfaces following final assembly.

The *E6* repeater is a plug-in unit and all external connections are made by means of a newly designed plug and connector. The connector has 16 dual-contact forked terminals spaced 1/8 inch apart in a molded block. The assembly is such that a floating action of each terminal is obtained. The terminal is made from phosphor bronze and is gold plated on the contact end. The wiring end is arranged for solderless wrapped connections. The over-all length of the connector is slightly less than 3 inches.

The plug has 16 gold-plated phosphor bronze contacts molded in a strip. These contacts engage the forks of the associated connector. The terminal ends of the contacts are arranged to be dip soldered to one of the printed wiring boards of the converter. In addition, provision is made for securing the plug with two screws to a separate guidance detail. This detail, which is an aluminum die casting, is designed to relieve the plug and connector of strain. When the repeater unit is plugged in on its mounting shelf, the tapered prongs of the guidance detail align the plug and the connector prior to engagement. A spring catch on the shelf engages one of the guidance prongs and locks the repeater in place.

The converter and the associated LBO

Vol.
Pt. 1
196

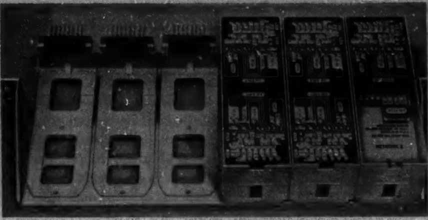


Fig. 24 (above). Shelf for six repeaters

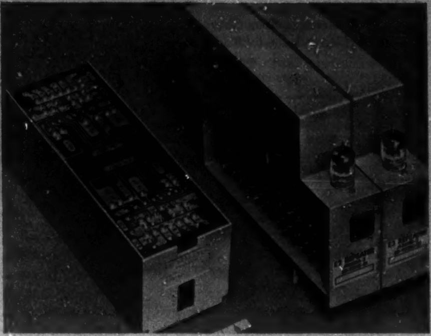


Fig. 25 (right). Comparison of E6 repeater with E2 and E3 repeaters

networks must, of course, be assembled into one complete working repeater unit. For this purpose an aluminum housing was designed; see Fig. 22. This consists of an H-shaped extrusion, smooth surfaces on the outside but provided with rails on the inner surfaces, and equipped with a small hinged door at one end. These rails match grooves in the phenolic blocks of the converter and LBO networks. The converter slides into position in the housing below the center rib, with the 16-contact plug facing the open end, or rear, of the unit.

The first LBO network slides into the housing above the center rib end from the front-door side. The second LBO network (assuming an intermediate repeater) slides into the housing above the center rib but from the opposite end. At approximately the middle of the housing there is a cutout in the center rib. Through this cutout the electric connections between the LBO networks and the converter are made by means of screws. When these connections are made, with the use of a molded spacer or filler block that drops into the center rib cutout, all the required parts making up a repeater unit are then interconnected and locked in place within the housing. For a terminal application of the repeater only one LBO network is required. In this case it is necessary to use a dummy network in place of the second LBO network, as shown in Fig. 23. Electrically, this dummy unit simply provides a wiring path from one side of the converter to one of the line connections.

The over-all dimensions of a complete repeater unit are $9^{19}/_{32}$ by $3^{3}/_{8}$ by $3^{9}/_{16}$ inches, and the weight of an intermediate repeater is $4^{1}/_{4}$ pounds.

In service the *E6* repeater unit is supported on shelves mounted on relay racks. The shelves are aluminum die castings made in two sizes, one for 19-inch racks holding five repeaters, and one for 23-inch racks holding six repeaters. These shelves, one of which is shown in Fig. 24, are so designed that the repeater is guided into position to engage the connector. The repeater is then automatically locked in place as described previously. As shown in Fig. 19, monitoring jacks of the installed repeater are accessible through a cutout in the door of the repeater housing. The jacks are actually mounted on the printed wiring boards of the converter. When it is desired to remove the repeater from service, the hinged door of the housing is opened and this automatically disengages the plug from the connector. The repeater unit may then be easily removed from the mounting shelf. The door action referred to here is obtained by a built-in cam in the hinged portion of the die-cast door.

A variety of relay-rack arrangements is available in order to satisfy various operating conditions. These details are beyond the scope of this paper; it is of interest to note, however, that the new design requires only half the space of the *E2-E3* electron-tube repeater. For a comparison, see Fig. 25.

Engineering

The fundamental difference between *E23* and the *E6* repeaters means that different techniques are used in the latter's application to the telephone plant. Both the design of an *E6*-repeatered circuit and its subsequent in-service line-up are different from the design and line-up of its predecessor.

The job of the circuit designer in using an *E6* repeater is twofold. He must first determine the effect the repeater will have on the circuit: the amount of available gain, the effect on signaling ranges, crosstalk and overload considerations, and the like. In short, he must determine if the repeater will do for the circuit what needs to be done. Second, he must determine the information the installation man will need to do an efficient line-up job. This in general involves giving the installer instructions for initial repeater adjustments so he will have a starting point for his line-up procedures.

One advantage of the *E6* repeater over the *E23* is that as an intermediate repeater it can be used between different types of facilities without sacrificing performance. For example, an *H-88* high-capacitance LBO network transforms line impedance to 900 ohms in series with 2 μf; a *D-88* high-capacitance network transforms an appreciably different line impedance to the same 900 ohms in series with 2 μf. The image impedance of the gain unit therefore can match the line in both directions. It should not be inferred, however, that the impedance of any type of cable facility can practicably be transformed to 900 ohms in series with 2 μf, but only line impedance not greatly different from 900 ohms. And only when line impedances on both sides of intermediate repeaters are transformed to the same value can the repeater match them both.

FACTORS CONTROLLING TOP GAIN

In the design of an *E6* repeatered circuit the allowable gain available from the repeater depends upon the return loss at the critical frequency, as seen by the repeater in both directions. These return losses are measures of reflections from the ends of the facility, structural imperfections, and irregularities in the intervening cable facilities, all referred to the location of the repeater. This is in line with conventional *E*-repeater design techniques where the repeater gain is

Table I. BOR Adjustment for Outside Cables of Uniform Gage

The 830A Network for H88 High-Capacitance Cable Terminal and Intermediate Repeaters

Resistance of Outside End Section Plus Office Cabling, Ohms	Resistance of BOR, Ohms	Screws Down
19-Gage Pairs		
0–42	28	1, 2; 1, 2
43–97	0	all
22-Gage Pairs		
0–40	84	1; 1
41–88	56	1, 3; 1, 3
89–127	28	1, 2; 1, 2
128–195	0	all
24-Gage Pairs		
0–15	196	none
16–46	168	3; 3
47–78	140	2; 2
79–109	112	2, 3; 2, 3
110–155	84	1; 1
156–187	56	1, 3; 1, 3
188–233	28	1, 2; 1, 2
234–311	0	all

Resistor screws must always be adjusted in pairs, viz., 1 and 1, 2 and 2, 3 and 3, never singly.

Table II. LBO Capacitor Adjustment Versus Equivalent End Section of Outside Cable

The 830A Network for H-88 High-Capacity Cable, Both Terminal and Intermediate Repeaters

End Section Length	19 Gage		22 Gage		24 Gage	
	BOC uf	Screws Down	BOC uf	Screws Down	BOC uf	Screws Down
Feet						
0	0.092	ACEFG	0.088	AEFG	0.081	DFG
200	0.089	BEFG	0.084	ABDFG	0.079	ACFG
400	0.086	ACDFG	0.081	DFG	0.077	ABFG
600	0.083	BDFG	0.078	CFG	0.074	FG
800	0.080	BCFG	0.075	AFG	0.072	ABDEG
1000	0.076	BFG	0.072	ABDEG	0.069	DEG
1200	0.073	CDEG	0.069	DEG	0.066	CEG
1400	0.070	ADEG	0.066	CEG	0.063	AEG
1600	0.066	CEG	0.063	AEG	0.060	CDG
1800	0.063	AEG	0.060	CDG	0.057	ADG
2000	0.060	CDG	0.057	ADG	0.054	ACG
2200	0.057	ADG	0.054	ACG	0.051	BG
2400	0.054	ACG	0.051	BG	0.048	ABDEF
2600	0.051	BG	0.048	ABDEF	0.046	ADEF
2800	0.048	ABDEF	0.045	DEF	0.043	ACEF
3000	0.045	DEF	0.042	CEF	0.040	BEF
3200	0.042	CEF	0.040	BEF	0.037	ACDF
3400	0.038	EF	0.037	ACDF	0.034	BDF
3600	0.035	ABDF	0.034	BDF	0.032	DF
3800	0.032	DF	0.031	BCF	0.029	CF
4000	0.028	ABF	0.028	ABF	0.026	AF
4200	0.025	F	0.025	F	0.023	ABDE
4400	0.022	BDE	0.022	BDE	0.020	DE
4600	0.019	BCE	0.018	ACE	0.017	CE
4800	0.016	ABE	0.015	BE	0.014	AE
5000	0.013	E	0.012	ACD	0.011	CD
5200	0.010	ABD	0.009	BD	0.008	AD
5400	0.007	D	0.006	BC	0.006	BC
5600	0.003	AB	0.003	AB	0.003	AB
5800 & Over	0	None	0	None	0	None

limited by the need to avoid singing and near-singing.

The terminal return losses used at the end of a trunk depend upon whether or not idle-circuit terminations or gain disablers are used. If they are not, a zero-db return loss (total reflection) is used since the circuit is open in the idle condition. If they are, a terminal return loss of 4.5 db is assumed at the originating end.

The structural return loss of the intervening cable facility depends upon the capacitance deviations in the cable itself, the inductance deviations in the loading coils, and the regularity of the load spacing. Preferably, the structural return losses used in designing an *E6* repeatered line would be measured values obtained in connection with cable completion tests. The use of such values permits obtaining the maximum possible gain from the repeater. In the absence of such data, estimates can be based on typical deviations of capacitance and inductance and on known load-spacing deviations.

It should be noted that the charts of allowable gain are based on return losses at the critical frequency, since the gain is limited by singing. The echo return losses used for terminal-link calculations in the Nationwide Switching Plan are not the same values, and they will lead, if used for this purpose, to improper repeater gains.

The terminal return losses are referred to the repeater location by adding twice the intervening loss. Of course, any lumped irregularity, such as a junction of two dissimilar facilities, a load-spacing irregularity of a piece of equipment, is also referred to the repeater. The various return losses are then combined on a power basis to obtain the return loss to be used in determining repeater gain.

One divergence in method arises here, because of the differences between the *E23* and the *E6* repeaters. The return losses upon which the repeater gains are based are those referred to the terminal of the *gain unit* through the LBO networks. Therefore, the losses of the LBO's are added to those of the cable facilities in referring the return losses. The losses of the LBO's, of course, depend upon how much building out they contain.

SETTINGS OF LBO NETWORKS

The LBO networks are designed to build out the impedance of the cable facility to 900 ohms in series with 2 μf so that it matches the repeater. As shown in Fig. 4, the LBO consists of an adjustable R and C line build-out, a fixed, HFC, and an adjustable LFC.

Adjustment of the R and C line build-out, of course, is related to length and gage of the cable end section to be built out. For engineering purposes, tables such as Tables I and II relate the length and resistance of the cable end section to the proper values of BOC and BOR.

The LFC is adjustable for gage only: 19, 22, or 24. For mixed gages adjustment is best made by trial during the line-up. For the rather usual case where finer-gage cable adjacent to the repeater does not extend out of the end section, guides are available with which the designer can specify approximate adjustments as line-up starting points.

The loss of the LBO is the sum of two parts: 0.3 to 0.5 db for the LFC, depending on gage; and the loss of a length of cable (simulated by the LBO) needed to build out the actual end section plus office cabling to 5,800 feet. Loss per unit length of the *loaded facility* is used in computing this latter part.

GAIN SELECTION

The allowable gain of the gain unit is shown on the chart in Fig. 26 as a function of the critical-frequency return losses referred to the repeater. Since this chart gives allowable gain of the gain unit by itself, the losses of the LBO networks must be included in determining the over-all losses of a circuit.

This limitation of gain given in the

allowable-gain chart is designed to provide adequate margin, in the order of 8 db in the average situation, against repeater-singing in service. This may seem excessive but it should be remembered that return losses representative of a group of circuits were used in the calculations, and that therefore many circuits will have margins considerably less than this.

The allowable gains of the *E6* repeater are higher than those of the *E23* for equal return-loss values by one db or more. High cable structural return losses are necessary to develop this additional gain. Representative structural return losses of at least 20 and preferably over 25 db at the critical frequency are necessary to exploit fully the additional gain capabilities of the *E6* repeater.

Crosstalk is generally a controlling factor only where the repeater is used at the end of a circuit, at which point the disturbing talker levels are high and the customer experiencing the crosstalk may be located quite close to the repeater. For the typical *22H-88* facility, maximum gains in the gain unit of about 7 db are recommended in the terminal situation. Intermediate repeater gains may of course be higher and are generally controlled by stability. Crosstalk coupling in the old nonstaggered twist cables, put in use around 1920 and before, is likely to be so high that repeaters with such cables are not recommended.

Effects on Signaling

The effect of the *E6* repeater on signaling must also be considered in the design of the circuit. The repeater not only passes but amplifies audio-frequency signals, such as single and multifrequency. It does, however, adversely affect the various low-frequency signals in different ways.

D-c supervision and 20-cycle ringing can be passed through the repeater without need for bypassing, subject, however, to the maximum loop-current limitation of 60 milliamperes. The increase in loop resistance caused by inserting the repeater depends upon whether it is a terminal or intermediate repeater, as well as upon the amount of building out used in the LBO network. An idea of the range of reductions involved may be gleaned from Table III which gives minimum and maximum effects.

In addition to the above loop-resistance effects, the shunt converter of the *E6* repeater bridges a capacitance across the pair. This has a distorting effect on pulse signals. The effect is quite complex and depends upon the length and resistance of the cable pairs and the type of central office and signaling systems used, as well as upon the adjustment of the relays in the signaling system.

Table III. Range Effects on Loop Signaling E6 Repeaters

	Loop Resistance, Ohms	
	Minimum Terminal E6, No Building-Out	Maximum Intermediate E6, Maximum Building-Out (Both Sides)
Supervision	65	482
20 cycles	400	792

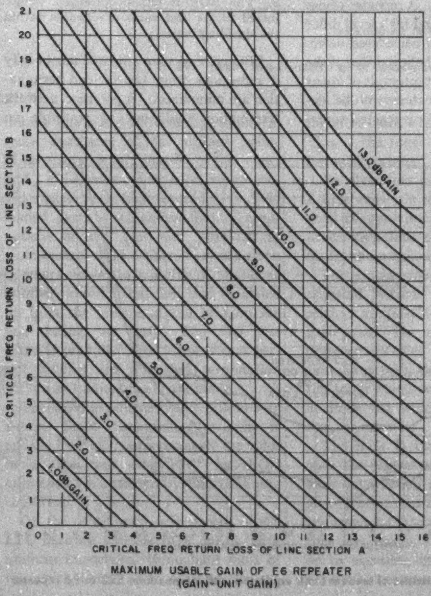


Fig. 26 (left). Maximum usable gain of E6 repeater (gain-unit gain)

In general, the distortion caused by the *E6* repeater is within operating margins for dial pulsing and revertive pulsing into crossbar offices and for panel-call-indicating (PCI) signals. Operation of certain revertive pulsing into panel offices and certain types of dial pulsing into step-by-step offices is adversely affected by the use of an *E6* repeater. The effect of the *E6* on pulsing has been summarized for the design engineer on a group of charts, which indicate certain restrictions that pulsing circuits impose on its use.

Example of Design

Let us take a typical example of an *E6*-repeatered trunk and go through a design calculation. Assume an 8-mile direct trunk using *24H-88* facilities as shown in Fig. 27. The trunk connects two offices with a signaling range of 2,400-ohm loop resistance. A terminal *E6* repeater is used. The representative stuctural return loss at the critical frequency has been found to be 26 db. No gain disablers or idle-circuit terminations are used.

The setting of the LBO are as follows: The 2,800-foot cable end section has a loop resistance of 145 ohms. From Table I the proper building-out resistor is found to be 84 ohms, obtained by screwing down the no. 1 screws. From Table II for a 2,800-foot end section of 24 gage, the proper BOC is found to be 0.043 obtained by turning down screws *A*, *C*, *E*, and *F*. The loss of the LBO is that of 5,800−2,800=3,000 feet of *24H-88* plus 0.5 db or 0.7+0.5=1.2 db.

The next step is to determine the return loss at the repeater. The terminal return loss is referred through the cable and LBO losses. The structural return loss is referred merely through

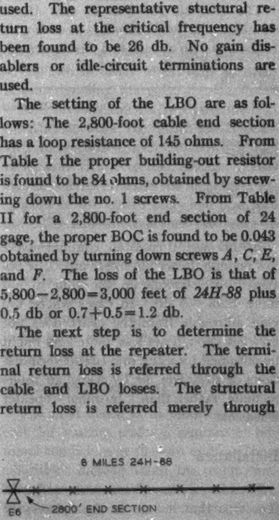


Fig. 27 (right). Typical example

the LBO loss. Both are then combined on a power basis.

$$\left.\begin{array}{l}\text{Terminal RL}\ .0+2(9.8)+2(1.2)\\ \qquad\qquad =22.0\\ \text{Structural RL}\ 26+2(1.2)\\ \qquad\qquad =28.4\end{array}\right\}=21\ \text{db}$$

where RL = return loss.

The central office side of the repeater, of course, presents a zero-db return loss, since no disabler or idle-circuit termination is assumed. The chart of maximum usable gains shows a gain-unit gain of 8.5 db for return losses of 21 db and zero db. Since this is a terminal repeater, crosstalk would normally limit this gain to about 7 db, although with modern unit-type 24-gage cable, the crosstalk characteristics might possibly permit the 8.5 db. Assume conservative designing, and use 7 db of gain.

For the benefit of the line-up man, 7 db of gain is specified, and if desired the proper screws to turn down can also be specified from a chart (in this case screws *C*, *E*, *J* and *2*, *4*, *5*, *7*, *8*).

Checking the signaling situation, add the loop resistance of the cable to the repeater resistance.

	Ohms
Cable, 8.0 miles, 24 gage	2,190
Gain unit of repeater	40
Building-out network (25 and 84)	109
Total	2,339

This meets the 2,400-ohm signaling range of the office.

The over-all net loss of this trunk should also include office loss at each end. If about 0.5 db at each end were assumed the over-all net loss would be as follows:

	Decibels
Cable facilities	9.8
Office loss	1.0
LBO	1.2
Total	12.0
Gain-unit gain	−7.0
Net loss	5.0

In this case, if crosstalk couplings were low enough, lower net losses might be obtained by stepping up the gain to not more than 8.5 db. In cases where top gain allowable from the chart does not give low enough net loss, more gain might be obtained by using idle-circuit terminations or disablers, or by using an intermediate, rather than a terminal, repeater if feasible.

Installation

Thus far, we have been concerned with the steps that led up to the delivery of the *E6* repeater at the place where it will be used. We known how it should perform in any given circuit; now, how will it deliver the expected performance?

Choice of LBO

A specific type of LBO network is needed for each type of cable loading. Initially, LBO networks are being produced for *H-88*-loaded and certain *D-88*-loaded facilities.

Adjustment of Gain and Image Impedance

Tables are available showing the screw adjustments needed to provide various gains. After the adjustments are made in accordance with the engineered gains, the gains of series and shunt elements of the gain unit are measured separately with the aid of the *54A* and *54B* sets as a check of image impedance. The image impedance of the two units together is correct if the gain of the series unit by itself equals the gain of the shunt unit by itself.

After the series converter and shunt-converter gains have been measured, and if necessary readjusted, the gain of the gain unit as a whole is measured and recorded on the front of the repeater as a reference from which drift can be computed.

Gain at higher than normal input level is measured using the *54A* set as a check of transistor condition. A repeater whose gains at normal level and high level differ by more than 0.4 db is considered defective. At present, it is believed that if the overload test is made at intervals of about a year, defective repeaters will be detected before they can have reacted appreciably on service. Experience may modify the present view of a desirable test interval.

Since the LBO network accomplishes the impedance match between cable facilities and the gain unit, its optimum adjustment is the essence of the high-grade performance of the *E6* repeater. Ideally, each LBO should be adjusted individually to the cable pair to which it is assigned. Practically, all LBO's in a group may be adjusted alike, with small sacrifice in performance if the group of facilities to which they are assigned is uniform and free from troubles.

The optimum adjustment of LBO networks is accomplished with the *54C* return-loss measuring set. This set consists of a sweep generator, a hybrid arrangement permitting direct optimizing of return loss, a transistorized detector, and controls. It is powered by 48-volt central office battery, but also uses 60-cycle power to control output frequencies through synchronous drive. Fig. 16 shows the sweep return-loss measuring set.

To start, the LBO networks are set at close to optimum adjustment, from knowledge of the type of facility and end section. Next, three separate adjustments are made in the following order:

1. The BOC is optimized.
2. The best LFC network is selected.
3. The BOR is either optimized or adjusted to the best value practicable within the signaling limits.

LBO networks are adjusted most easily if reflections from the far-end termination are minimized. A precision network termination built out to resemble an

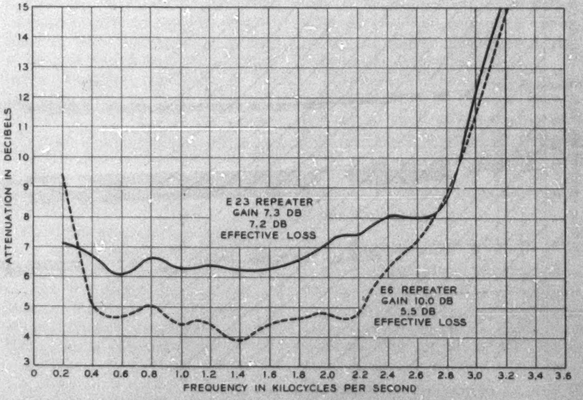
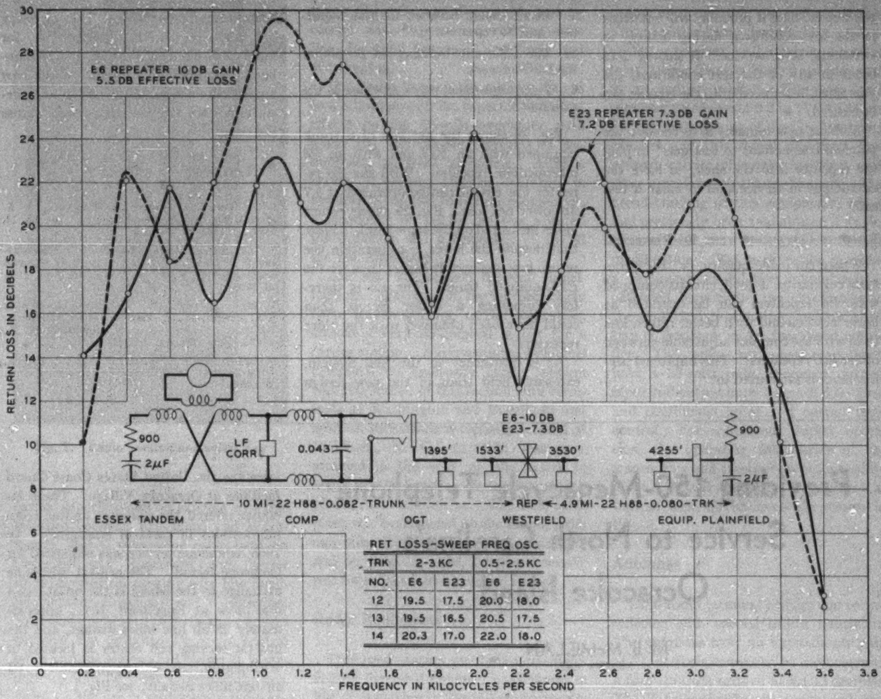


Fig. 28. Net loss characteristic of tandem trunk equipped with intermediate E23 or E6 repeater



RET LOSS-SWEEP FREQ OSC

TRK	2-3 KC		0.5-2.5 KC	
NO.	E6	E23	E6	E23
12	19.5	17.5	20.0	18.0
13	19.5	16.5	20.5	17.5
14	20.3	17.0	22.0	18.0

Fig. 29. Return loss characteristic of tandem trunk equipped with intermediate E23 or E6 repeater

infinite line would be ideal, but would require assistance at the distant point. If necessary, a termination of 900 ohms and 2 μf in series can be used, which provides moderately high return loss to *H-88* loaded pairs of 0.5 end section. Terminations of this kind are provided in most central offices, where they are given numbers similar to subscribers' telephone numbers in order that they may be dialed up from the other end and held on the line during the tests. The *54B* test stand is provided with features that permit holding up a dialed termination while LBO adjustments are made.

The BOC is adjusted with a 2- to 3-kc sweep. Since the singing margin is usually least at the upper end of the voice band, the best sensitivity for BOC adjustment is obtained by thus confining the sweep.

On the other hand, adjustment of the LFC network and the BOR are of fairly uniform importance throughout most of the voice band. The adjustments are therefore made with a 0.5- to 2.5-kc sweep. This is the range in which echo return loss is important.

To save time in optimizing, the BOC is first varied in steps of about 0.004 μf until an approximate maximum is reached. Finer variations may then be made to attain the maximum.

For single-gage trunks, the LFC is adjusted without test, since standard settings are provided for 19-, 22-, and 24-gage pairs. For mixed gages, the circuit designer specifies a setting from rules that have been determined empirically; the tester starts with this setting, but if he finds one which gives higher return loss, he uses it.

The BOR is then adjusted. Starting with the suggested resistance, the tester finds out whether a lower or higher value gives greater return loss. He always uses the resistor that provides the highest return loss except when an upper limit is set by signaling limitations. In case two different resistance settings give the same high return loss, he uses the lower of the two in order to provide all practicable signaling margin.

After all adjustments have been made in the LBO network, the return loss is measured and recorded, both with 0.5- to 2.5-kc sweep and with 2- to 3-kc sweep. These measurements indicate the structural excellence of the loaded pairs and how well the LBO network matches the pairs to the gain unit. They are useful for reference in case of cable trouble.

If the repeater is an intermediate one, the LBO network on the other side of the repeater must be adjusted. This is done in the same way as the first, but entirely independently.

Monitoring Check and Test Call

The final check is to monitor the circuit for singing under a number of terminal conditions at the two ends. These conditions include about the worst combinations that could occur at any time. If no singing occurs under any of these conditions, then it probably will not occur in the idle condition, during dialing, in switching, or in service. If singing does occur in any of the test conditions, the gain must be reduced until the trouble can be cleared.

All that now remains is to remove the patched connections to test gear, to plug the repeater into the shelf, to have the circuit put in service, and to make a test call.

Results Obtained with E6 Repeater

Where the cable quality, i.e., the structural regularity, is good, circuits equipped with *E6* repeaters can be operated at lower net loss and with better return loss than with the complex adjustable network of the *E23* repeaters. This improved performance is attributed to:

1. Better match between the line impedance and *E6* repeater.
2. The adjustable building-out portion of the LBO network.
3. Finer adjustment steps provided in the gain unit.

Fig. 28 shows the net loss of a trunk equipped with an *E23* and with an *E6* intermediate repeater. With the *E6* repeater the circuit can be operated at 1.7 db lower net loss. Fig. 29 shows return losses measured on the same trunk. Note that despite the lower loss operation the return loss is generally higher than for the *E23*-equipped trunk. The *E6* is therefore providing a better circuit than could have been obtained with the older repeaters.

As is customary in the Bell System, extensive field trials of the new design have been undertaken. The circuitry has been on trial in Atlanta, Ga., for nearly 3 years. In addition, a large number of preproduction models of the new repeater are now in actual use on commercial circuits in a number of locations across the country.

References

1. A Negative Impedance Repeater, J. L. Merrill, Jr. *AIEE Transactions*, vol. 69, pt. II, 1950, pp. 1461–66.

2. Transistor Negative-Impedance Converters, J. G. Linvill. *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 41, June 1953, pp. 725–29.

3. Negative Impedance Telephone Repeaters, J. L. Merrill, Jr., A. F. Rose, J. O. Smethurst. *Bell System Technical Journal*, New York, N. Y., vol. 33, Sept. 1954, pp. 1055–92.

4. Semiconductor Signal Translating Device, S. Darlington. *U. S. Patent no. 2,663,806*, Dec. 22, 1953.

Providing 150-Megacycle Telephone Service to North Carolina's Ocracoke Island

M. B. McMILLAN
MEMBER AIEE

EXTENDING along the east coast of North Carolina from Manteo in the north to Atlantic in the south lies a chain of islands known as the Outer Banks. Within this chain are Hatteras, Ocracoke, and Portsmouth. It is with the island of Ocracoke that this paper is concerned.

The island of Ocracoke is approximately 16 miles long and has an average width of less than one mile. On the north, it is bounded by Pamlico Sound and Hatteras Inlet; on the east, by the Atlantic Ocean; on the south, by ocean and sound; and on the west, by the sound only. The island may be reached from the north by automobile over paved roads on Hatteras and Ocracoke Islands and by State-operated ferries across Oregon and Hatteras Inlets. Mail boat service operating daily between Atlantic and Ocracoke offer passenger service from the south, and private plane service is available from several mainland cities. The island is inhabited by approximately 850 residents, the greater number of whom live in Ocracoke Village at the southern end of the island. The village is adjacent to the inlet and is built around a salt-water lake. The pursuits of the sea occupy most of the island's 850 population.

The only communications prior to 1955 were the the United States Coast Guard facilities at Ocracoke Village. These facilities offered the islanders limited communications in cases of emergency. In 1955, telephone service was requested for Ocracoke Island. The nearest telephone exchange to the island is the unattended dial office at Engelhard, N.C., approximately 28 air-line miles distant, and the nearest serving toll center is located in Washington, N. C., approximately 97 air-line miles distant; see Fig. 1.

Surveys of the families and business establishments on the island indicated that facilities which would provide a total of three communications channels for off-the-island communications and that a 100-line dial office located at Ocracoke Village should be sufficient to satisfy demands for service. Comparative cost studies were made utilizing: 1. radio facilities between Ocracoke and Washington, 2. radio facilities between Ocracoke and Engelhard and

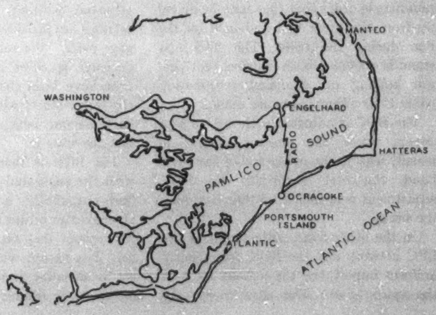


Fig. 1. Map showing location of Ocracoke Island, N. C.

Paper 60-793, recommended by the AIEE Radio Communication Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting, Atlantic City, N. J., June 19–24, 1960. Manuscript submitted March 21, 1960; made available for printing April 26, 1960.

M. B. McMillan is with the Carolina Telephone and Telegraph Company, Tarboro, N. C.

Fig. 2. Radio transmitter-receiver bay

carrier superimposed on existing wire line facilities between Engelhard and Washington, and 3. submarine cable between Ocracoke and Engelhard and carrier superimposed on existing wire line facilities between Engelhard and Washington. This study indicated that the three trunks between the proposed Ocracoke central office and the serving toll center at Washington could be most economically provided by the use of multiplexed radio between Ocracoke and Engelhard, and by carrier superimposed on existing wire line facilities between Engelhard and Washington.

The desire to keep the antenna height to a minimum and the expected reflection effects on the 28-mile overwater path made the 152–162-mc (megacycle) frequencies the most attractive. After a survey of existing stations operating in the area in the 152- to 162-mc frequencies, a frequency of 152.57 mc was selected for Ocracoke and a frequency of 157.83 mc for Engelhard.

The wire line section between Engelhard and Washington, having previously been transposed for carrier operation, offered a very economical layout for a carrier system. A conventional 4-channel compandored open-wire carrier system was established in this section with terminals at Washington and Engelhard and with an intermediate repeater located at Belhaven, N. C. A room was constructed in the rear of the hotel to house the 100-line dial office and the multiplexed radio equipment. The local cable plant was constructed on the island, radio and wire line facilities were connected, and on June 15, 1956, telephone service was established to Ocracoke Island.

Radio Equipment

Packaged station combinations modified for extended bandwidth, as shown in Fig. 2, operating in the 150-mc frequency range, were used for the radio facilities. This combination included a transmitter, a receiver, and a carrier-operated relay, tropicalization-treated and mounted in an indoor cabinet. The base-band frequency response of 0.2 to 11.2 kc allows for the multiplexing of a voice-frequency channel and two single side-band suppressed carrier channels, together with three associated signaling channels. The broad-banding of this equipment, which had previously been available on a single-channel basis, was accomplished through modification. The modification provides a frequency-modulation characteristic, extended bandwidth, the elimination of the peak deviation limiter, and the addition of feedback in the base-band amplifier. The units are phase-modulated, have a 50-watt radio-frequency output, input and output impedances of 600 ohms, and continuous input and output level control. For continuous duplex operation, the operating frequencies of the transmitter and receiver are separated by 5 mc. Actual operating frequencies as well as transmitted levels are shown in Fig. 3.

Antennas

Each radio terminal utilizes one corner reflector with special iridite treatment. The antennas have an approximate gain of 7.5 decibels compared to a dipole with a front-to-back ratio of about 15 decibels. The Engelhard reflector is mounted on a

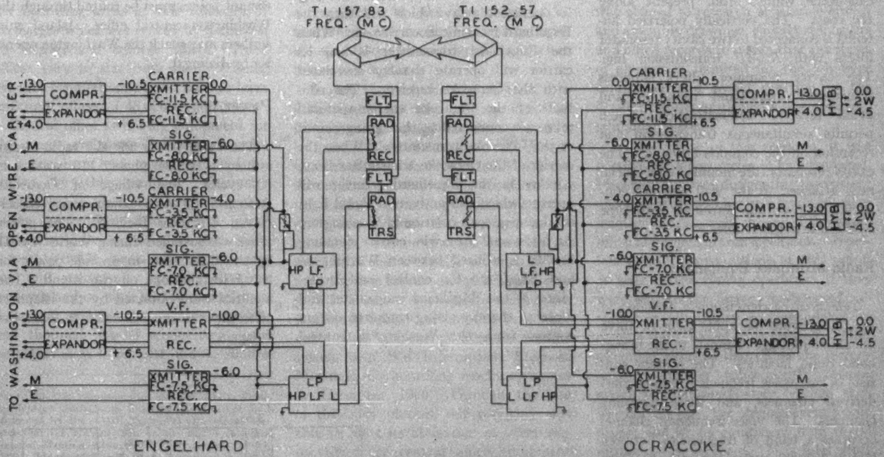

T1 157.83 FREQ. (MC)
T1 152.57 FREQ. (MC)
TO WASHINGTON VIA OPEN WIRE CARRIER
COMPR. EXPANDOR
CARRIER XMITTER FC-11.5 KC REC. FC-11.5 KC
SIG. XMITTER FC-8.0 KC REC. FC-8.0 KC
CARRIER XMITTER FC-3.5 KC REC. FC-3.5 KC
SIG. XMITTER FC-7.0 KC REC. FC-7.0 KC
V.F. XMITTER REC.
SIG. XMITTER FC-7.5 KC REC. FC-7.5 KC
-13.0 +4.0 -10.5 +6.5 0.0 -6.0 -4.0 -10.0
M E
FLT. RAD REC FLT. RAD TRS
L HP LF LP
LP HP LF L
L LF HP LP
LP L LF HP
HYB. 0.0 2W -4.5
ENGELHARD
OCRACOKE

Fig. 3. Block diagram of radio-carrier system with transmission levels. All levels in decibels referred to one milliwatt with 1,000-cycle tone at zero transmission level point

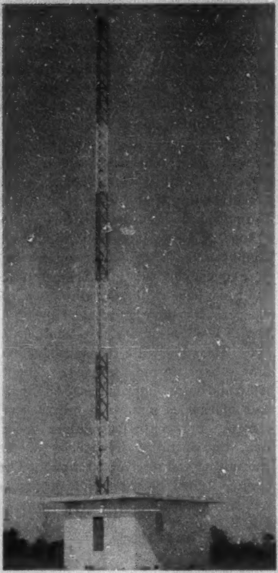


Fig. 4. Engelhard dial central office building and antenna structure

150-inch guyed tower as shown in Fig. 4. The Ocracoke Island reflector is mounted on a 65-inch pole. Reinforced guying is used for protection against the hurricane-force winds, often present along the coast. The vertically polarized antenna is connected to two cavity resonator filters with *RG-8U* transmission line. The cavity resonator filters suppress transmitter extra-band radiation and avoid receiver desensitization, which permits simultaneous transmission and reception on the common antenna. The cavity resonators are mounted on a special rack attached to the wall and are connected to the radio terminals with type *S-450* transmission lines.

Radio Multiplex Equipment

Telephone carrier channels, as shown in Fig. 5, provide for application of three telephone message channels to the broadband radio facilities. The system consists of one voice frequency, two carrier-derived channels, and three signaling channels. The voice-frequency channel occupies a band of frequencies between 200 and 3,000 cycles per second. Two additional amplitude-modulated single-side-band suppressed carrier channels are derived by modulation of voice frequencies of 300 to 2,700 cycles per second. Channel 2 transmits and receives on the upper side band of 3.5 kc and channel 3 transmits and receives on the lower side band of 11.5 kc, as shown in Fig. 6. The block diagram of three channels of the radio carrier system, including compandors and voice-frequency terminating units, are shown in Fig. 3. Signaling for dialing and supervision on the three channels are transmitted and received at carrier frequencies of 7.5, 7.0, and 8.0 kc for channels 1, 2, and 3 respectively. Each channel utilizes a compandor unit to provide an effective increase in signal-to-noise ratio.

Radio System Alarms and Controls

The alarms and controls of the radio system appear as both audible and visual alarms on the Washington, N.C., central office switchboard with instructions to authorized personnel for their interpretation and operation. A red pilot light indicates when the carrier is on or off. Red light on indicates that the carrier is on, while red light off indicates that the carrier is off. In addition to this visual alarm, an audible alarm will sound to indicate that the carrier is off.

The method used to turn the transmitter carriers on and off is as follows: The contacts of a switch located on the Washington switchboard are extended over an open-wire carrier-signaling facility to Engelhard. These facilities are used to operate a relay which will turn the Engelhard transmitter carrier on. When the Engelhard transmitter is on, its carrier will operate a relay associated with the Ocracoke receiver. The contacts of the Ocracoke carrier-operated relay are connected to the keying circuit of the Ocracoke transmitter. When the carrier of the Ocracoke transmitter is on, a relay in the Engelhard receiver will operate which will activate a signal light on the operator's position in Washington. An additional open-wire carrier signaling facility is utilized between Washington and Engelhard for control and supervision of the Engelhard transmitter and receiver, thereby giving complete system control from the Washington control point. The carrier system is so arranged that failure of the open-wire facility on which the carrier system is superimposed will result in a shutdown of the radio system.

Fig. 5. Radio multiplex equipment bay

Dial Central Office

The new office located at Ocracoke Village is a step-by-step community dial central office with a capacity of 200 individual lines. The office is arranged for normal 2-letter 5-digit dialing and is identified as WAverly 8. While local calls are made on the 2-letter 5-digit basis, incoming and outgoing calls to distant points must be routed through the Washington central office. Island subscribers may reach the Washington operator by dialing 0.

Power

The basic power for the radio-carrier system is single phase, 115 volts a-c, 60 cycles. The village of Ocracoke generates its own power, using diesel-driven generator facilities. The power plant was installed during World War II by the U.S. Navy to provide power for Navy installations on the island. The facilities were obtained by the islanders after the war and expanded to serve the village.

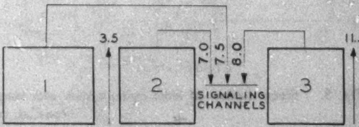


Fig. 6. Radio multiplex channel allocations

A privately owned ground return power system serves the town of Engelhard. This system operates without a neutral conductor and uses the earth for the return path. Emergency a-c power for the Engelhard office is provided by diesel-driven automatically controlled alternators with a single-phase 60-cycle output of 7.5 kw at 240/120 volts a-c. In the event of a power failure, the generator will start automatically and will continue to operate until the power is restored.

Conclusions

The use of the radio-carrier system in conjunction with the open-wire carrier system and the dial office has provided the isolated community on Ocracoke Island with a modern, reliable telephone service. The acceptance of this service by the islanders is evidenced by the fact that 50 additional lines have been installed since the initial installation was made.

References

1. One and Three Channel Packaged VHF Radio Systems. *Bell System Practices*, New York, N. Y., sect. R104.050, issue 1, June 1959.

2. One and Three Channel Packaged VHF and UHF Radio Systems for Point-to-Point Applications. *Ibid.*, issue 2, Aug. 1959.

3. Voice Frequency and Carrier Telephone System. Lenkurt Electric Company, Inc., San Carlos, Calif., form 33B1-14939-DES, issue 1, Mar. 1955.

4. Fifty Watt High Band Station Combinations with Accessory RC-208. *Instruction Manual EBI-4990*, General Electric Company, Schenectady, N. Y., issue 1, Sept. 1955.

Automation in Switching Equipment Engineering

M. E. HACKER
ASSOCIATE MEMBER AIEE

C. M. DAVIS
NONMEMBER AIEE

L. 79
1
60

THE General Telephone Company of California is presently the largest operating company within the General Telephone System. Its major operating territories embrace the Los Angeles metropolitan area and include important exchange areas situated in other sections of central and southern California. For many years, this company has standardized on the use of step-by-step switching equipment manufactured by the Automatic Electric Company. Included in our operating areas at this time are 37 exchanges which are served by a total of 70 step-by-step central offices. These offices vary in size from small terminal per-line installations of 200 lines to large multiunit offices of 43,000 terminals.

Central office switching equipment for the many exchanges furnishing service to subscribers was engineered and installed by this company. Company engineering to meet switching equipment needs was established in order to insure that our individual requirements with respect to traffic, special switching apparatus and community growth changes, however unique, were met with minimum delay. It was felt that a more complete and thorough job of engineering could be achieved at lower cost and on shorter time intervals by our own engineers than would be possible by using conventional factory engineering services.

The period of accelerated growth experienced by the company since World War II made substantial changes in earlier engineering methods imperative. As reflected in Fig. 1, the General Telephone Company of California is presently serving in excess of 1,000,000 stations. This quantity represents an increase of approximately 540% since 1945, and was accomplished by the establishment of many new central offices as well as numerous additions to existing offices to meet the service demand.

During the early part of this era, detailed equipment lists and specifications for switching equipment were prepared by hand. Written data from an engineer were transcribed to typewritten sheets for subsequent delivery to the supplier. As the need for switching equipment grew, it became evident that equipment engineering personnel would have to be augmented by a significant number to meet the then current and predicted future growth needs. Appropriate mechanization studies instituted at this time were culminated by a change from the manual system to an IBM (International Business Machines Corporation) machine process in December 1956. Through the use of this new process, our company was able to meet its obligation of effectively providing for central office equipment growth requirements. At the same time, it was possible to accomplish this task with fewer engineers than would have been required under the previous method of operation. Not only were the engineering processes made measurably more economical, but greater job uniformity and a significant reduction in engineering and clerical errors resulted.

Manually Prepared Equipment Lists

Subsequent to the receipt of proper authority for engineering additions to automatic telephone switching systems, or new systems, the detail engineer reviewed the project traffic order to determine equipment types, quantities, layout, trunking, power, and all related requirements. The engineer's first order of business was to start immediately an equipment order for the factory, to minimize any delay in receipt of equipment due to manufacturing processes. Before actually preparing this order, however, the engineer was obliged to research applicable circuit and assembly drawings, closed work orders and allied records to insure that the items of equipment required for the job were compatible and would function in the desired manner as elements of a complete switching system.

In the course of this research, it was possible for inaccuracies to be introduced. Piecepart numbers designating the variety of components making up a given job were lengthy and, as such, enhanced the possibility of numerical transpositions. Needless to say, transpositions at this point, or later in the typing operation, if they were sent to the factory undetected, resulted in unnecessary correspondence to resolve the problem thus created. This, of course, assumes that either the factory or the engineer subsequently became aware of the discrepancy. When such was not the case, the obvious event would occur; the wrong equipment item would ultimately be received. It is interesting to note at this point that numerical transpositions in a piecepart number often represent a valid part number, further

---

Paper 60-778, recommended by the AIEE Communication Switching Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting, Atlantic City, N. J., June 19–24, 1960. Manuscript submitted November 2, 1959; made available for printing March 28, 1960.

M. E. Hacker and C. M. Davis are with the General Telephone Company of California, Santa Monica, Calif.

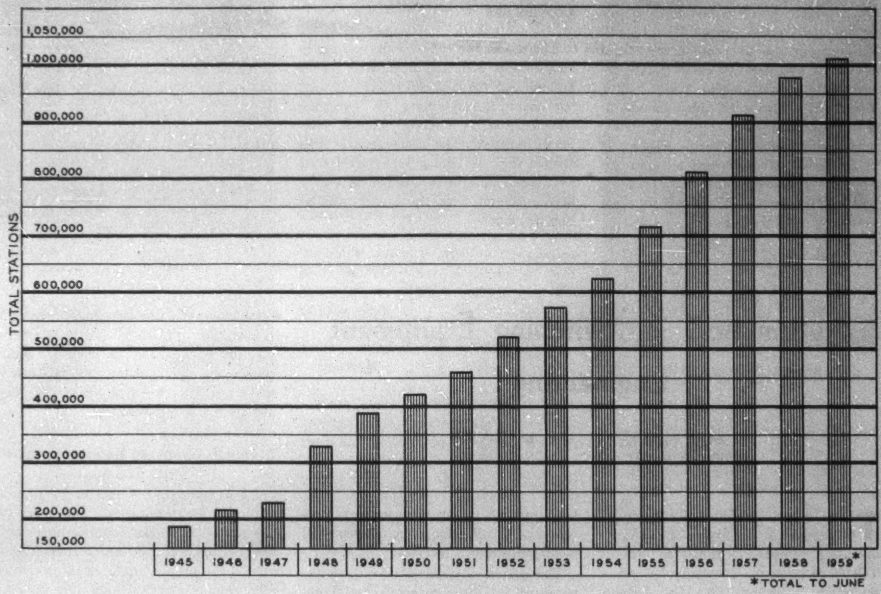


Fig. 1. Graph of station growth since World War II

complicating any error detection procedure.

METHOD OF PREPARATION

When sufficient data had been accumulated for the project, the engineer would originate a handwritten order preparatory to typing. This order was prepared on a special group of forms containing all necessary facts pertinent to the engineering project at hand. As noted in Fig. 2, a switch count summary was originated to reflect the general types and quantities of switching equipment required for the job. Specific details covering each individual item to be ordered from the factory were then written on other forms; see Fig. 3.

Following a general description of the use of the equipment listed in Fig. 3, the engineer would continue by listing the item number, quantity required, manufacturer's part number, and a brief word description of the first item of equipment required for the job. Where applicable, circuit or assembly numbers were included in the description to minimize manufacturing errors. Thereafter, subsequent items were listed in a similar manner. As noted in the figure, all equipment required for the job was grouped as a matter of convenience. Thus, for the project of Fig. 3, distributing frame equipment became group 1; trunk board equipment, group 2; etc. Through this procedure, manufacturing as well as engineering processes were greatly facilitated.

In conjunction with this operation, the detail engineer was required to obtain current equipment prices in order to allow a reasonably accurate estimate to be made relative to the items to be purchased. These cost figures were not to be included on the factory equipment list, but were accumulated primarily for work order and budget purposes to cover the installation of the switching equipment.

PREPARATION TIME AND REPRODUCTION CONSIDERATIONS

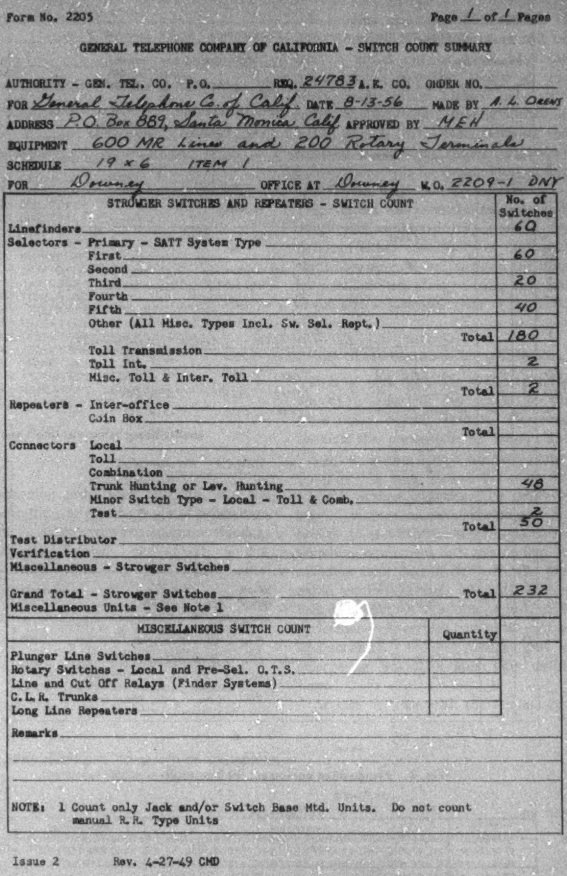
The amount of time required for the preparation of an equipment order on the basis just described varied in proportion to the complexity and size of a given switching system. Accordingly, situations requiring the provision of directors, automatic toll ticketing, and related switching apparatus involved considerably more detail engineering time for the preparation of an equipment list than did a job associated with the establishment of, or an addition to, a small local step-by-step central office. For these reasons, equipment lists for the factory might consist of as few as one or two pages or as many as 80 or more. Engineering time for the completion of a handwritten list could therefore vary from a matter of hours to as much as 6 to 8 weeks and, on occasion, even longer.

Throughout this period of manual preparation of the equipment list, the problem of transcription or clerical errors was ever present. Diligent effort on the part of the detail engineer, the typist, and supervisory personnel reduced to a practical minimum, but could not entirely eliminate, the difficulties attributed to this source. During this era, a constant stream of correspondence between the factory and the company strove to rectify the various errors detected and to clarify any deviations from standard factory ordering procedures.

It should be mentioned at this point that other clerical activities concerned with the typing of the completed pencil equipment list also contributed to the general problems associated with the manually prepared equipment list engineering method. Due to the number of copies required of each list, the legibility of some carbon copies of each sheet was impaired. Additionally, since this equipment list subsequently became a part of an installer's specification, it was necessary to type the same list twice; once to order the equipment and a second time months later when the installer's instructions were prepared for distribution. Aside from the handicap of semi-illegible copies, the necessity of typing each equipment list twice unnecessarily increased engineering costs.

Summary

From the foregoing general description, it is apparent that the typing routine, along with numerous other procedures, required attention in order to place equipment engineering on a more efficient basis as the company grew. The monotonous, routine task of transcribing ordering information and shop notes on the more lengthy equipment lists accounted for our initial interest in mechanizing such work. Removing this tedious clerical type of work from the shoulders of the detail engineer would free him, in a minimum time, to accept new engineering projects. Toward this end, studies were made to determine the feasibility of utilizing automatic business machines for the production of complete equipment specifications. The details of this effort are outlined in the paragraphs to follow.



Form No. 2205 Page 1 of 1 Pages

GENERAL TELEPHONE COMPANY OF CALIFORNIA - SWITCH COUNT SUMMARY

AUTHORITY - GEN. TEL. CO. P.O. ______ REQ. 24783 A.E. CO. ORDER NO. ______
FOR General Telephone Co. of Calif. DATE 8-13-56 MADE BY A. L. Owens
ADDRESS P.O. Box 889, Santa Monica Calif. APPROVED BY MEH
EQUIPMENT 600 MR Lines and 200 Rotary Terminals
SCHEDULE 19 x 6 ITEM 1
FOR Downey OFFICE AT Downey W.O. 2209-1 DNY

STROWGER SWITCHES AND REPEATERS - SWITCH COUNT			No. of Switches
Linefinders			60
Selectors -	Primary - SATT System Type		
	First		60
	Second		
	Third		20
	Fourth		
	Fifth		40
	Other (All Misc. Types Incl. Sw. Sel. Rept.)		
		Total	180
	Toll Transmission		
	Toll Int.		2
	Misc. Toll & Inter. Toll		
		Total	2
Repeaters -	Inter-office		
	Coin Box		
		Total	
Connectors	Local		
	Toll		
	Combination		
	Trunk Hunting or Lev. Hunting		48
	Minor Switch Type - Local - Toll & Comb.		
	Test		2
		Total	50
Test Distributor			
Verification			
Miscellaneous - Strowger Switches			
Grand Total - Strowger Switches		Total	232
Miscellaneous Units - See Note 1			

MISCELLANEOUS SWITCH COUNT	Quantity
Plunger Line Switches	
Rotary Switches - Local and Pre-Sel. O.T.S.	
Line and Cut Off Relays (Finder Systems)	
C.L.R. Trunks	
Long Line Repeaters	

Remarks ______

NOTE: 1 Count only Jack and/or Switch Base Mtd. Units. Do not count manual R.R. Type Units

Issue 2 Rev. 4-27-49 CMD

Fig. 2. Handwritten switch count summary

Machine-Prepared Equipment Lists

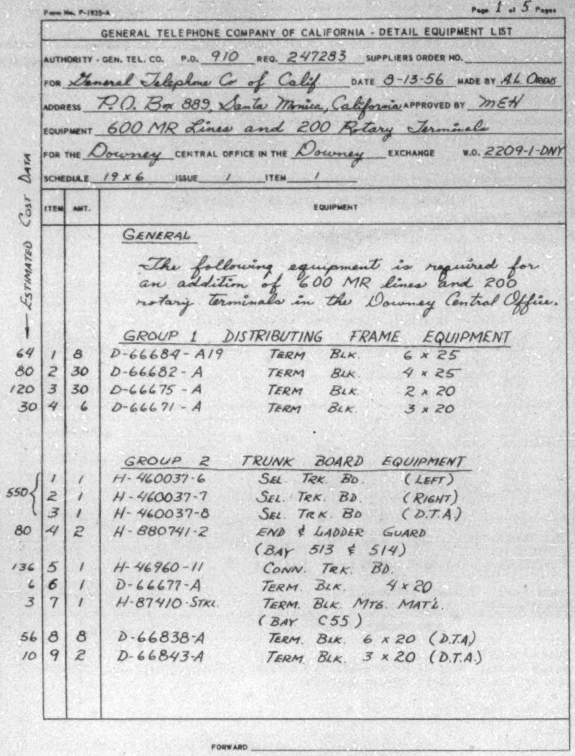
Investigations leading to the use of automatic business machines for the production of switching apparatus equipment lists were based primarily on three major objectives, as follows:

1. To increase the productivity of the detail engineer.
2. To reduce the need for additional clerical and engineering personnel during a period of accelerated company growth.
3. To minimize the "human errors" present in the manually prepared equipment list system.

Prior to the actual implementation of such a program, it was necessary to devise a standard, uniform method with which to identify the myriad components comprising an automatic telephone system. Therefore, to make the preparation of equipment lists with IBM equipment a practical reality, such items as linefinders, selectors, connectors, director equipment, earthquake bracing, terminal blocks, cable, and miscellaneous mounting hardware had to be provided with their own unique identification numbers.

Subsequent to the compilation of a complete listing of all required equipment items, the systematic development of a flexible identification scheme was approached on the basis of segregating the basic switching system elements into fundamental "building blocks of equipment." Thus, from such building blocks an engineer would find it possible to order complete or partial switching systems merely by specifying only those items required for a given project. Accordingly, a numerical coding system was evolved to identify all items in each building block or order specification as they came to be known. From these listings, the detail engineer would be able to convey conveniently to an IBM key punch operator both equipment-type and quantity information from which appropriate control cards could be prepared.

To insure that the over-all objectives of machine-prepared equipment lists would be satisfied to the greatest extent possible, close co-ordination of various developmental activities was maintained between our company and the switching equipment manufacturer. At the inception of our initial studies, certain factory procedures



Form No. P-1935-A

Page 1 of 5 Pages

GENERAL TELEPHONE COMPANY OF CALIFORNIA - DETAIL EQUIPMENT LIST

AUTHORITY - GEN. TEL. CO. P.O. 910 REQ. 247283 SUPPLIERS ORDER NO.

FOR General Telephone Co of Calif DATE 8-13-56 MADE BY A.L. Orcus

ADDRESS P.O. Box 889, Santa Monica, California APPROVED BY MEH

EQUIPMENT 600 MR Lines and 200 Rotary Terminals

FOR THE Downey CENTRAL OFFICE IN THE Downey EXCHANGE W.O. 2209-1-DNY

SCHEDULE 19 x 6 ISSUE 1 ITEM 1

← ESTIMATED COST DATA	ITEM	AMT.	EQUIPMENT
			GENERAL
			The following equipment is required for an addition of 600 MR lines and 200 rotary terminals in the Downey Central Office.
			GROUP 1 DISTRIBUTING FRAME EQUIPMENT
64	1	8	D-66684-A19 TERM BLK. 6 x 25
80	2	30	D-66682-A TERM BLK. 4 x 25
120	3	30	D-66675-A TERM BLK. 2 x 20
30	4	6	D-66671-A TERM BLK. 3 x 20
			GROUP 2 TRUNK BOARD EQUIPMENT
550	1	1	H-460037-6 SEL. TRK. BD. (LEFT)
	2	1	H-460037-7 SEL. TRK. BD. (RIGHT)
	3	1	H-460037-8 SEL. TRK. BD (D.T.A.)
80	4	2	H-880741-2 END & LADDER GUARD (BAY 513 & 514)
136	5	1	H-46960-11 CONN. TRK. BD.
6	6	1	D-66677-A TERM. BLK. 4 x 20
3	7	1	H-87410-STKL. TERM. BLK. MTG. MAT'L. (BAY C55)
56	8	8	D-66838-A TERM. BLK. 6 x 20 (D.T.A.)
10	9	2	D-66843-A TERM. BLK. 3 x 20 (D.T.A.)

FORWARD

Fig. 3. Handwritten equipment list face sheet

SPEC. NO.	DESCRIPTION
582	(MANUAL) PABX ATTENDANT'S CABINET, PBX SWITCHBOARDS, SMALL C. B. SWITCHBOARDS, SPECIAL MANUAL BOARDS NOT CLASSIFIABLE IN OTHER MANUAL SPECS. (AND WHEN OTHER MANUAL SPECS. ARE NOT USED, RELAY EQUIPMENT AND RELAY RACKS ASSOCIATED WITH SPECIAL MANUAL BOARDS).
592	(AUTOMATIC) TELETYPEWRITER AND MISCELLANEOUS EQUIPMENT NOT CLASSIFIABLE IN OTHER SPECS.
606	TOLL BOARD (CORD OR AUTOMATIC), RURAL BOARDS, AUXILIARY (D.S.A.) SWITCHBOARDS, ETC.
702	CABLE (FORMED AND BULK) AND JUMPER WIRE.
703	RUNWAYS, ERECTION MATERIAL AND HARDWARE (CHANNELS, WALL ANGLES, ETC.) LIGHTING AND CONDUIT (ILLUMINATION EQUIPMENT) AND ROLLING LADDER EQUIPMENT.
722	(AUTOMATIC) EQUIPMENT FRAMES (IRONWORK) AND UNASSEMBLED RELAY RACKS (USE 746 FOR MANUAL).
723	AUTOMATIC SWITCHING EQUIPMENT (LINE EQUIPMENT, LINE FINDERS, SELECTORS, CONNECTORS, REPEATERS AND TRUNK CIRCUITS) AND SUBSCRIBER'S STATION EQUIPMENT.

Fig. 4. Partial list of Automatic Electric Company job specifications

were reviewed to determine that the proposed IBM program would conform to manufacturing needs wherever possible. At the same time, an existing factory "job specification" identification system was adopted for use by our company in setting up and maintaining an effective machine ordering process. Fig. 4 reveals a partial list of these job specifications.

It might be helpful at this point to differentiate between "order" and "job specifications." Simply stated, an "order spec" is, as its name implies, a means for the engineer to order the equipment required to satisfy the needs of his engineering project. A given project may require the use of one or several order specs, depending upon the complexity of the system. A completed order spec is sent to the IBM group for machine production of a final equipment order for the factory. The resulting order is then known as an "equipment list" and is ready to be forwarded to the factory together with an appropriate purchase order. The equipment list consists of one or more job specs which conform exactly to factory spec numbers. In other words, order specs, which are used by the telephone company engineer are converted by the IBM process into job specs as required by the factory in their manufacturing procedures.

Upon the inauguration of our IBM program, routines were established to control all equipment-ordering activities utilizing the IBM process. This included the addition of new order specs when required, inserting new items into existing specs and revising equipment groupings for greater efficiency where necessary. In this manner, the detail engineer was assured of having up-to-date standard equipment items from which to design efficient telephone switching systems.

NEED FOR ORDER SPECIFICATIONS

Order specs were created primarily to facilitate the application of business machine processes to the needs of switching equipment engineering. Indeed, they were to become the basic ingredient in the ordering process, and as such every effort was made to furnish them in a form which was as complete and flexible as conditions would permit. Any feature which was considered to be of value in reducing the actual time required by an engineer to prepare a complete order was included. As implied by Fig. 4, a definite need for order specs existed because factory job specs included, in most cases, far too great a variety of switching apparatus under one specification number. A single IBM order spec based on a given job spec would have contained so many pages that it would have been of little practical value to either the detail engineer or IBM key punch operator.

To derive the greatest benefits from the IBM process, therefore, arrangements were made to provide a more detailed breakdown of the equipment manufacturer's job specifications. Fig. 5 illus-

ORDER SPECIFICATION CODE BLOCKS

000-099	UNASSIGNED	300-399	UNASSIGNED	600-699	PABX EQUIPMENT
100-199	DIAL EQUIPMENT	400-499	UNASSIGNED	700-799	SAGE
200-299	DIAL EQUIPMENT	500-599	UNASSIGNED	800-899	UNASSIGNED

SPEC. NO.	DESCRIPTION	SPEC. NO.	DESCRIPTION
100	TRK BDS & RLY RKS	126	DIESEL
104	TERM BLKS	128	POWER BOARD - 1000 AMP
105	TRK BDS & RLY RKS - MANUAL	129	SDSP & ACSP EQUIP
106	ROLLING LADDERS	130	MISC POWER
107	CABLE & WIRE	131	MISC POWER
110	PRI SUPER MATL	132	MISC POWER
111	SEC SUPER MATL	135	MISC LIGHTING
113	SEC SUPER SUPP EARTHQUAKE BR ASSEM	136	MISC LIGHTING
114	SEC SUPER SUPP - ROD SUPP ASSEM	137	MISC LIGHTING
115	CABLE RACK MATL	167	TEST TURRET
116	SEC TOP ANGIES FOR EQUIP BAYS	183	OAF & SAF
117	MISC CHANNEL, BAR & ANGLE IRON	185	TONE & INT
118	SUPER DETAILS	201	LINEFINDERS
119	SUPER & CA RACK FITTINGS	202	SELECTORS - 1ST TO 5TH & TOLL ROUTE
120	BATTERIES - ELECTRIC STORAGE	203	SELECTORS - TOLL TRANS
121	POWER BOARDS & COPPER BUS BAR	204	SELECTORS - TOLL INT
122	RING MACH EQUIP	206	CONNECTORS
123	BATTERY CHARGERS - LORAIN	207	REPEATERS - MTD
124	BDFB EQUIP - SPLIT BUS	208	REPEATERS - LOOSE SHELVES & SWS
125	BDFB EQUIP - SINGLE BUS	210	MISC BAY EQUIP

Fig. 5. Partial index of General Telephone Company order specifications

trates a portion of the order specs which were established to meet our engineering needs from the factory job specs. In the course of planning for a practical system of ordering information, it became clear that in some cases even further subdivision was required for the convenience of the detail engineer. An example of this may be seen in Fig. 5, where selector equipment is concerned. A single order spec for this equipment would have consisted of approximately 84 pages. By assigning order specs 202, 203, and 204 according to specific types of selectors, a single, large specification was avoided as described below:

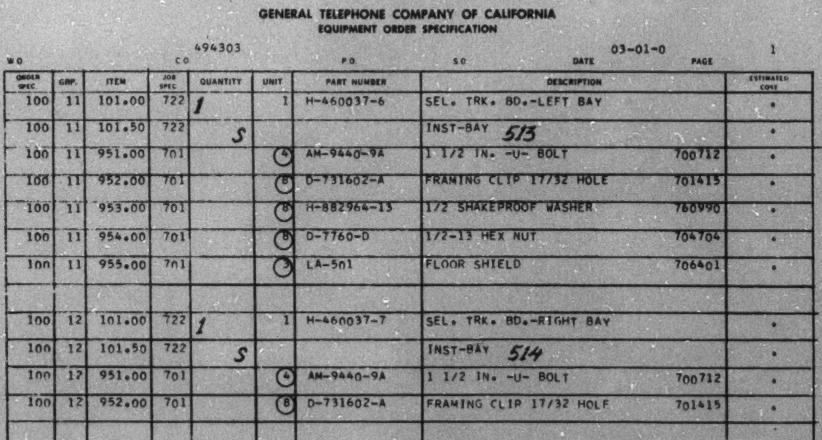
Specification No.	Title	No. of Pages
202	Selector equipment, first, fifth, and toll route	43
203	Toll transmission selector equipment	18
204	Toll intermediate selector equipment	23

Depending upon the specific needs of an engineering project, selectors will be ordered under one or more of these specifications. For this reason an engineer need concern himself only with such spec sheets as are directly applicable to his project. This clearly avoids the necessity of searching through irrelevant data and accounts for a significant saving in the detail engineer's productive effort. This concept was extended to include all comparable equipment types which, by virtue of either complexity or numerous physical arrangements, dictated the need for a further reduction to basic elements to conform to the objectives of the IBM program.

Order Specification Coding System

It is beyond the scope of this paper to present a complete description of the order spec coding system. However, through the paragraphs to follow, and the accompanying figures, a general understanding will be gained of the more important system features. As will be readily understood, the mechanized approach to preparing complete telephone switching equipment lists was no less demanding than that of the manually prepared equipment lists. Indeed, in the face of a machine incapable of judgment, performance directives had to be explicit and accurate if the end product were to emerge in an acceptable and usable form.

As was the case with the manual sys-

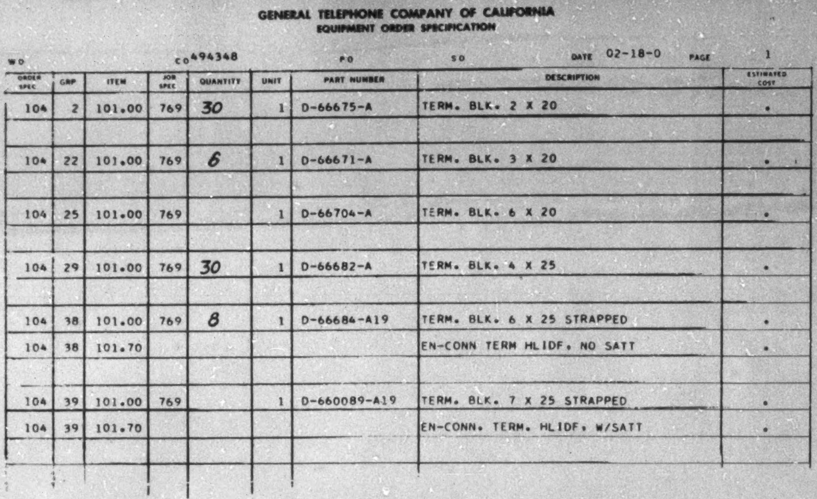


GENERAL TELEPHONE COMPANY OF CALIFORNIA
EQUIPMENT ORDER SPECIFICATION

W O
C O 494303
P O
S O
DATE 03-01-0
PAGE 1

ORDER SPEC	GRP.	ITEM	JOB SPEC	QUANTITY	UNIT	PART NUMBER	DESCRIPTION	ESTIMATED COST
100	11	101.00	722	1	1	H-460037-6	SEL. TRK. BD.-LEFT BAY	.
100	11	101.50	722	S			INST-BAY 513	.
100	11	951.00	701		④	AM-9440-9A	1 1/2 IN. -U- BOLT 700712	.
100	11	952.00	701		⑧	D-731602-A	FRAMING CLIP 17/32 HOLE 701415	.
100	11	953.00	701		⑧	H-882964-15	1/2 SHAKEPROOF WASHER 760990	.
100	11	954.00	701		⑧	D-7760-D	1/2-13 HEX NUT 704704	.
100	11	955.00	701		③	LA-501	FLOOR SHIELD 706401	.
100	12	101.00	722	1	1	H-460037-7	SEL. TRK. BD.-RIGHT BAY	.
100	12	101.50	722	S			INST-BAY 514	.
100	12	951.00	701		④	AM-9440-9A	1 1/2 IN. -U- BOLT 700712	.
100	12	952.00	701		⑧	D-731602-A	FRAMING CLIP 17/32 HOLF 701415	.

Fig. 6. Sample of 100 order spec



GENERAL TELEPHONE COMPANY OF CALIFORNIA
EQUIPMENT ORDER SPECIFICATION

WO CO494348 PO SO DATE 02-18-0 PAGE 1

ORDER SPEC	GRP	ITEM	JOB SPEC	QUANTITY	UNIT	PART NUMBER	DESCRIPTION	ESTIMATED COST
104	2	101.00	769	30	1	D-66675-A	TERM. BLK. 2 X 20	.
104	22	101.00	769	6	1	D-66671-A	TERM. BLK. 3 X 20	.
104	25	101.00	769		1	D-66704-A	TERM. BLK. 6 X 20	.
104	29	101.00	769	30	1	D-66682-A	TERM. BLK. 4 X 25	.
104	38	101.00	769	8	1	D-66684-A19	TERM. BLK. 6 X 25 STRAPPED	.
104	38	101.70					EN-CONN TERM HLIDF. NO SATT	.
104	39	101.00	769		1	D-660089-A19	TERM. BLK. 7 X 25 STRAPPED	.
104	39	101.70					EN-CONN. TERM. HLIDF. W/SATT	.

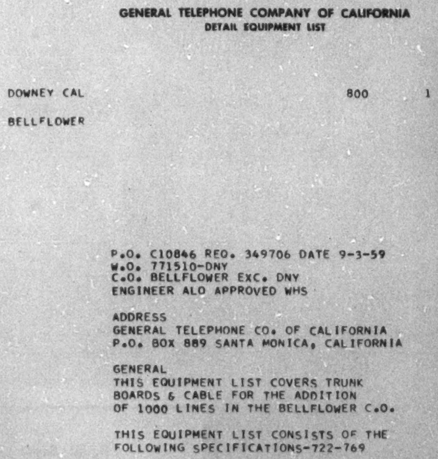
Fig. 7. Sample of 104 order spec

tem, automated equipment lists required appropriate item, group, quantity, piece-part number, description, and similar information. At the same time, a straightforward stationery form was required to facilitate its completion by the detail engineer. Referring to Figs. 6 and 7, it will be seen that all of the above requirements were satisfactorily met through a coding system and form arrangement which became standard for all order specs. At present, there are 108 order specs available for equipment engineering purposes with a combined total of approximately 1,258 pages. These specifications encompass some 20,000 individual equipment items and, in addition, provide for the acquisition of specialized items where required by a need unique to a given engineering project.

Fig. 5 illustrates the 3-digit code system utilized in the order spec assignment scheme. Additionally, this figure contains a listing of the ten code number blocks into which various equipment engineering functions are divided. It will be noted that ample provision has been made for the expansion of the IBM program to include any future engineering requirement which lends itself to the adaptation of automatic business machine processes.

Group 1 of each order spec was reserved for an index of the various groups which were to make up an individual specification. This index served to aid engineering processes since, if the engineer required but a few items in a particular specification, rapid access to the items of interest was assured. Subsequent groups were listed in a numerical sequence to the maximum required by each basic specification. Each group within an order spec will become an "item" in the final equipment list.

The item number was perhaps the most significant coding unit developed for the order spec system. Basically it consists of a 5-digit-plus-decimal-point code and serves a wide variety of purposes. Its primary objective is to identify each of the various lines on an order spec. In addition, it satisfies several other equally important requirements. Aside from the item or spec line number itself, the 5-digit-plus-decimal-point code provides for the identification of the supplier from whom an item will be purchased, identi-



GENERAL TELEPHONE COMPANY OF CALIFORNIA
DETAIL EQUIPMENT LIST

DOWNEY CAL 800 1

BELLFLOWER

P.O. C10846 REQ. 349706 DATE 9-3-59
W.O. 771510-DNY
C.O. BELLFLOWER EXC. DNY
ENGINEER ALO APPROVED WHS

ADDRESS
GENERAL TELEPHONE CO. OF CALIFORNIA
P.O. BOX 889 SANTA MONICA, CALIFORNIA

GENERAL
THIS EQUIPMENT LIST COVERS TRUNK
BOARDS & CABLE FOR THE ADDITION
OF 1000 LINES IN THE BELLFLOWER C.O.

THIS EQUIPMENT LIST CONSISTS OF THE
FOLLOWING SPECIFICATIONS-722-769

Fig. 8. Typical equipment list face sheet

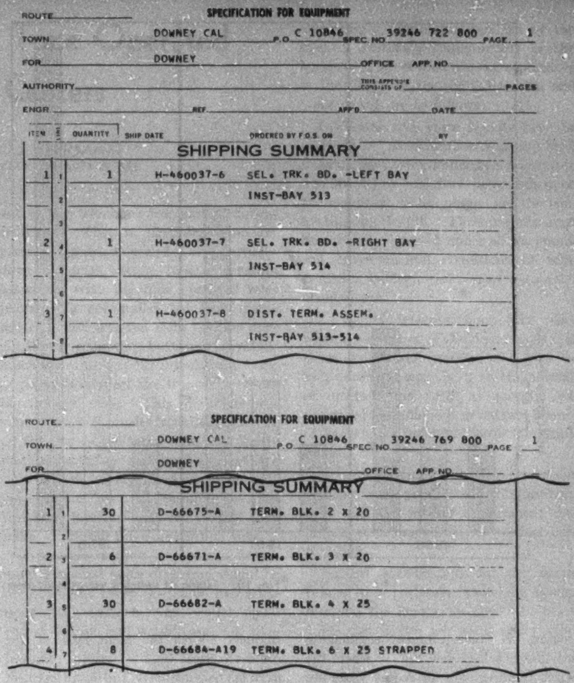
fies all installer's, engineer's and shop notes in a uniform manner, and signals the IBM machine whenever tabulation is required to accumulate totals of similar types of equipment such as nuts, bolts, washers, etc. Substantial experience in using the various order specs has proved that adequate study and planning during the formative stages of our IBM program has resulted in an item-coding scheme which inherently contains the maximum in utility and flexibility.

Mentioned briefly in the manually prepared equipment list description were two other items related to arithmetic computations; these were the switch count summary and the addition of estimated equipment costs. As might be expected, the IBM equipment utilized for the production of switching equipment lists is ideally suited for these operations. These data are conveniently available as by-products of the IBM process and they effectively remove additional clerical activity from the duties of the detail engineer. Routines are provided whereby up-to-date cost information is regularly inserted into the master IBM card file for each order spec.

As each order spec was completed, during the developmental phases of the mechanization program, it was thoroughly checked and subsequently transferred, line by line, to a master IBM card file. From this master group of cards, printed copies of each order spec were then available upon which could be entered data pertinent to an engineering project. The properly notated form was then converted by a key punch operator into control cards which were subsequently used by a tabulating machine operator to pull the required groups of cards within an order spec from within the master card file.

The cards thus selected are then reproduced with the copies serving as a basis for processing the current engineering job. Through the use of a master deck of IBM cards for each order spec, the amount of key punching is kept to a minimum, since the master deck is reassembled into its original form after duplicate cards are prepared for each job.

Upon the completion of the various card-sorting, deleting, arithmetic and other machine functions involving the duplicate cards, a final printed equipment list is available for the factory. At this point any number of copies of the equipment list may be obtained with each copy as equally distinct as the original. This feature alone represents another significant advantage over the manually prepared equipment list method.



SPECIFICATION FOR EQUIPMENT

ROUTE

TOWN DOWNEY CAL P.O. C 10846 SPEC. NO. 39246 722 800 PAGE 1

FOR DOWNEY OFFICE APP. NO.

AUTHORITY THIS APPENDIX CONSISTS OF PAGES

ENGR. REF. APP'D. DATE

SHIP DATE ORDERED BY F.O.S. ON BY

SHIPPING SUMMARY

ITEM	LINE	QUANTITY		
1	1	1	H-460037-6	SEL. TRK. BD. -LEFT BAY
	2			INST-BAY 513
	3			
2	4	1	H-460037-7	SEL. TRK. BD. -RIGHT BAY
	5			INST-BAY 514
	6			
3	7	1	H-460037-8	DIST. TERM. ASSEM.
	8			INST-BAY 513-514

SPECIFICATION FOR EQUIPMENT

ROUTE

TOWN DOWNEY CAL P.O. C 10846 SPEC. NO. 39246 769 800 PAGE 1

FOR DOWNEY OFFICE APP. NO.

SHIPPING SUMMARY

1	1	30	D-66675-A	TERM. BLK. 2 X 20
	2			
2	3	6	D-66671-A	TERM. BLK. 3 X 20
	4			
3	5	30	D-66682-A	TERM. BLK. 4 X 25
	6			
4	7	8	D-66684-A19	TERM. BLK. 6 X 25 STRAPPED

Fig. 9. Typical equipment list detail sheets

While necessarily brief, the above description suggests the achievement of the desired goals established at the beginning of the investigations concerning the use of business machines to relieve the detail engineer of unnecessary clerical activities. The order spec system is the key to this success, and studies continue for improved methods of more fully utilizing the various features available in the IBM equipment. Figs. 6 through 9 depict a before and after situation concerning a typical order for switching equipment. Figs. 6 and 7 represent a completed order spec as the IBM key punch operator receives it, and Figs. 8 and 9 are examples of a completed equipment list as received by the factory for manufacturing purposes. The estimated cost column in Figs. 6 and 7 has been left blank, although in an actual list it would be filled in by the IBM process.

Another item covered under the manual preparation method involved specifications for the equipment installer's use. By design, the complete standardization of the mechanized ordering process allows the use of the same job spec, which is forwarded to the factory, as the basis of an installers specification. With the addition of the relatively few instructions peculiar to an individual job, a complete specification is then available which covers all phases of the installation of the switching system equipment.

Conclusions

Our company, as a segment of the independent telephone industry, recognized the various problems to be faced during the postwar years. With local communities growing "by leaps and bounds" it became abundantly clear that a change was necessary in our equipment engineering methods to keep pace with the ever-increasing demand for telephone switching facilities. An exploratory program utilizing automatic business machines for the production of certain switching equipment engineering requirements was instituted. The success and subsequent effect of this

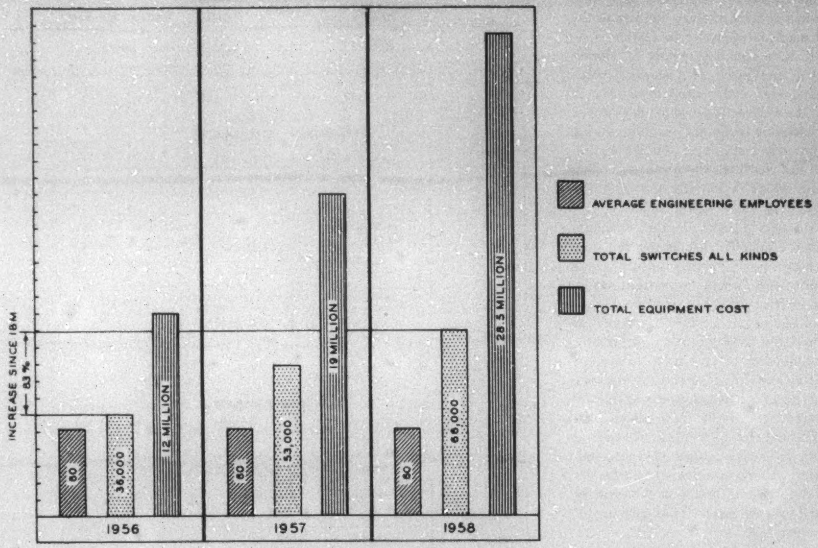


Fig. 10. Graph of switches versus manpower

program on our equipment engineering activities is reflected in Fig. 10 where the switch requirements for recent years are plotted against the number of detail engineers required to handle the work load adequately. It is evident from this chart that the efficiency of the detail engineers was increased substantially by the mechanization of the ordering process. Not only were greater numbers of engineering projects completed during a given length of time, but it was possible by the adoption of the IBM process to attain this goal with little increase in the total complement of engineering personnel.

Our mechanization program has accomplished the three primary objectives outlined earlier in this paper. Through the use of automatic business machines we have increased the production capability of each engineer, minimized the need to hire additional engineering employees during a period of rapid growth, and achieved a remarkable reduction of all types of "errors" on switching equipment lists.

It is interesting to note also that certain other less tangible benefits have been forthcoming with the change to machine-prepared equipment lists. One advantage was in the form of improved job interest on the part of the individual engineer. With the great bulk of the routine clerical type of activity removed from his responsibility, the detail engineer could devote a greater portion of his time to the more interesting phases of switching system design. Not only was the morale of the individual improved in this case, but the telephone company benefited by an improvement in the quality of the projects being completed. When considered collectively, the several advantages accruing from the use of automatic business machines as a tool for equipment engineering purposes actually exceed those originally contemplated.

The present IBM system for equipment-ordering purposes represents but one facet of a long-range plan, which implies further important business machine uses in the future. An example of this is found in current studies which concern themselves with the adaptation of engineering reference drawings to IBM equipment. Because of the nature of our present manually prepared card index for reference drawings, it is not possible to furnish copies to the detail engineer. It is anticipated that a "mechanized" approach to this need will allow the publication of master and supplemental indexes to meet any engineering reference requirement. Convenient access to engineering drawings, either by equipment type, operating features, or drawing number, will contribute to a further increase in the efficiency of our equipment engineering force.

Our equipment engineering automation efforts utilized IBM punched card equipment merely as a matter of convenience, since a substantial installation of this equipment existed within the company. It is recognized that any number of equally effective machine processes could doubtless produce the same results and, perhaps, do so in a more efficient manner.

Needless to say, opportunities exist for the application of mechanized processes ahead of as well as behind the one described in this paper. It is entirely within the realm of possibility for the present typewritten traffic specification to be reduced to punched card data which, together with appropriate directives, could result in a more fully integrated equipment engineering operation. Along the same line, the output of our present system might be advantageously modified to furnish punched cards or tape to the manufacturer in lieu of the present printed specification. Such a procedure would allow the Automatic Electric Company to use this data directly for manufacturing processes and avoid the need to translate a verbal equipment list into a machine input form at the factory.

Voice Immunity of Tone Operated Switching Centers

J. H. GUENTHER
NONMEMBER AIEE

J. SHIRMAN
ASSOCIATE MEMBER AIEE

TONE operation of electronic switching centers is the natural outgrowth of 4-wire operation of communication systems using a radio link. The radio circuit may be anywhere between the subscriber and the switching center or between switching centers.

Once the decision to use tone operated devices had been reached, the next major decision was to determine what part of the spectrum should be utilized for the control tones. For ease of design the obvious answer would be to use a frequency band for control which is separate from the information frequency band. A little reflection shows the fallacy of this thinking. The bandwidth requirement becomes out of proportion to its use. The switching function operates during such a small period of the total circuit holding time that to utilize an expanded bandwidth solely for this function is wasteful in an already crowded frequency spectrum. A separate signal band is also expensive from a components point of view as separating filters and signaling circuitry are needed at every switching stage.

For these reasons in-band supervisory and keying tones which are within the frequency spectrum required for the transmission of intelligence were chosen as the mode of operation. The use of in-band tones presented one major problem. How to prevent voice energy from operating the switching circuits?

In a telephone system all switching functions except release are normally performed before the flow of any information. This leads to an obvious solution to the problem; that is, provide a lock out of the telephone transmitter during switching. This is partially done on present impulse dials and is also partially done on push-button tone generators associated with the tone operated telephone. To accomplish this the transmitter is disconnected from the line during the time the tone button is operated, much the same way the impulse dial back contacts disconnect the transmitter during the dial cycle. This operation prevents voice frequencies from being present on the line at the same time that legitimate supervisory and keying tones are present. However, this does not prevent voice frequencies from operating the control tone detectors between control tone pulses or during the conversation time.

A possible method to accomplish immunity against voice operation would be to provide at every telephone two additional tone buttons. One would signal end of dialing, and would cause the tone detectors at the switchboard to lock out before the telephone is connected to the line. The other button would signify end of call and would unlock detectors before completion of the call. This method is not acceptable for two reasons. First is the high cost involved. Any cost added to the telephone set proper is of course multiplied by the number of telephones used. The second reason is the unaccustomed operation required. Additional operations required of the subscriber would result in increased false calls and condemnation of the switching system. The most acceptable solution to the problem is to provide voice immunity in the tone detection circuits at the switching center.

The system here described was designed under Signal Corps Contract No. *DA-36-039-SC-72356* and *DA-36-039-SC-72790* for the Army field switchboard. Operating specifications were set by the over-all switching requirements of the system. The specifications of importance in the design of the digit detection and voice guard are as follows:

1. Tones and voice signals may be received at the switchboard at any level from 0 dbm (power in decibels referred to 1 milliwatt) to −30 dbm, depending upon the condition of the phone and the connecting line length. These signals must be correctly detected under all the standard environmental conditions of temperature, battery, component parameter changes, etc.

2. Signal detection must take place when the duration of the signal tones is as short as 50 ms (milliseconds) and as long as several seconds; however, a break of up to 8 ms such as may be experienced by radio fade or contact bounce of keysenders must not record multiple digits.

3. These signals must be received and detected in the presence of 10 db (decibels) S/N with the further requirement that normal voice tones shall not produce false signaling.

Further specifications for the operation of this particular switching system reserved the frequency spectrum below 1,500 cps (cycles per second) for other digital processes. This necessitated setting the tone frequencies at the upper portion of the spectrum. Specifically, eight tones from 1,700 cps to 2,900 cps at 200-cps intervals were selected for control purposes.

Laboratory investigation along with research into the available literature led to the decision that each supervisory and keying signal should be composed of a 2-frequency tone. Accordingly, all supervisory functions except "seize" are performed by a 2-frequency compound tone. It was considered unnecessary to protect against false seize initiated by voice due to the nature of telephone operation, that is, the instant the phone is taken off the hook seize tone is sent out. Any voice tones that could simulate seize will have no effect because that tone is present anyway. As soon as seize tone is detected and dial tone is returned by the switching

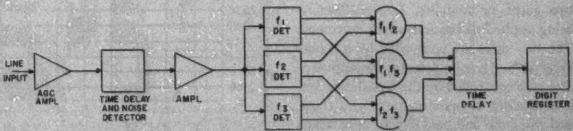


Fig. 1. Block diagram of basic digit detection circuit

Paper 60-820, recommended by the AIEE Communication Switching Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting, Atlantic City, N. J., June 19–24, 1960. Manuscript submitted March 17, 1960; made available for printing April 27, 1960.

J. H. GUENTHER and J. SHIRMAN are with the Stromberg-Carlson Company, Rochester, N. Y.

The authors acknowledge the many contributions made by the engineers of the Telecommunication Department, Stromberg-Carlson Company. Acknowledgments are also extended to the members of the Switching Branch, Data Processing Division, Communications Department, U. S. Armed Services Research an dDevelopment Laboratories, for their assistance during the development phases of this project.

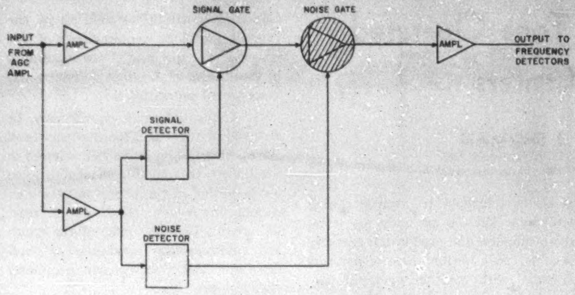


Fig. 2. Block diagram of noise detector circuit

center the seize detectors are removed from the circuit and any further seize tone on that line will have no effect on the seize detectors. The signal detection circuits were designed for voice immunity simultaneously with the design for operational requirements.

Fig. 1 is a logic block diagram of the basic signal detection circuits. For clarity only 3-frequency detectors are shown.[1-3] Signal reception and detection is accomplished by amplifying the line input with an automatic-gain-control (AGC) amplifier which produces an output of 0 dbm for any input signal level in the range of 0 dbm to −30 dbm; and by then detecting the individual control tones by single-frequency-pass-filters followed by storage circuits. Voice guard action is provided by time delay and noise detection circuits.

Before designing the time delay and noise detection circuits it was necessary to determine the frequency components and their relative strength and duration in the noise spectrum. In this connection two types of noise must be considered; white noise such as might be encountered in a high gain radio link, and voice during signaling. White noise and voice do not have the same energy distribution with frequency.

As white noise has a relatively even distribution of energy over the entire frequency spectrum a circuit for rejecting white noise is easily accomplished. Fig. 2 shows such a circuit. The signal from the AGC amplifier is made to pass through two gates, or switches, in series before being applied to the frequency detectors. The two gates are controlled by a signal detector and a noise detector respectively. The gate controlled by the signal detector is equivalent to open contacts of a relay and will close and pass signals only when the signal detector operates. The gate controlled by the noise detector is just the opposite, that is, equivalent to closed contacts of a relay and opens to prevent an output when the noise detector operates.

Due to the even energy-frequency character of white noise such a circuit is relatively simple in design. The signal detector is comprised of a bandpass filter, tuned to pass the signal frequencies of 1,700 through 2,900 cycles. This filter is followed by a trigger circuit which produces a d-c output when energized with a pulsed input. Thus the presence of any frequency or frequencies in the signal band will operate the signal gate.

The noise detector is similar but opposite in action. A band reject filter, designed to complement the signal filter, will pass any or all frequencies which fall outside the signal band. This filter is followed by another trigger circuit which produces a d-c output of opposite polarity. This d-c signal will be removed in the presence of noise energy and thus inhibit signals from passing through the noise gate.

By adjusting input attenuators in the signal and noise detectors it is possible to make the noise gate operate at a 10 db S/N. For this part of the specification "noise" is defined as "white noise."

Because of the operating characteristics of the AGC amplifier it is necessary to provide a time delay of approximately 10 ms (milliseconds) minimum before either the noise detector or signal detector operates. This minimum time allows the output of the AGC amplifier to stabilize. The first few cycles of AGC output are highly distorted and, if allowed to pass on to the detector circuits, the distortion products would operate the noise gate circuit and inhibit the signal.

This basic circuit fulfills the requirement of detection in the presence of 10 db S/N, but further modification of design was necessary to provide for voice immunity. Before designing for voice immunity it was necessary to establish the energy distribution versus frequency of the voice. Reference 1 provided valuable information about the average voice characteristics. Fig. 3 is a sound spectrogram of the words "speech we may see," reproduced by permission of the publisher and authors of "Visible Speech," by Potter, Kopp, and Green.[1] This sound spectrogram has time as a horizontal axis, frequency as a vertical axis, and intensity recorded as density, using a 300-cycle bandwidth analyzer. This is representative of voice characteristics.

Analysis of the instantaneous speech sounds reveals differences which are contrary to the generally accepted ideas of speech energy distribution. Classical curves showing speech energy distribution against frequency are based on long-time averages, and indicate maximum power output in the vicinity of 400 cps with very little average power in the 1,700- to 2,800-cycle range. However, analysis of sound spectrogram patterns for various English words and phrases show many areas of high intensity sound in the 1,700 to 2,900 cps band for considerable time durations; that is considerable compared with a maximum 50-ms detection time. Sounds such as ch, th, ss, and ee have energy peaks which appear almost entirely within the 1,700–2,900-cps signaling band. These sounds are responsible for talkdown of the basic circuit as described and necessitated a design approach to eliminate, or at least reduce, their effect.

Referring again to Fig. 1, special time delay circuits were designed into two different areas of the circuit to provide this voice immunity. Fig. 4 shows the

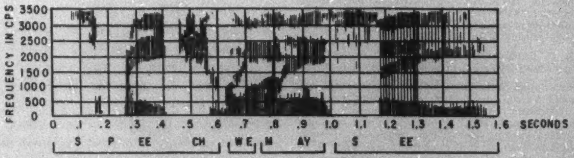


Fig. 3. Sound spectrogram of the words "speech we may see"

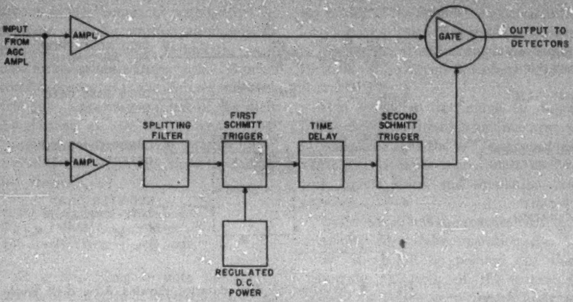


Fig. 4. Block diagram of control detector

block arrangement of the control circuit for either signal or noise gate. The output of the AGC amplifier is separated in the splitting filter and applied to a first Schmitt trigger circuit. The purpose of this trigger circuit is to provide constant amplitude pulses to the time delay circuit. Battery variations are removed by the regulator circuit which changes the battery variation from approximately 20% to 2%. The time delay circuit provides for a controlled delay for the operation of the second Schmitt trigger and for a fast release at the end of the signal. This is especially important in the noise detector as a pulse of noise should not block the detector for an indefinite length of time.

Fig. 5 shows the basic arrangement of the time delay circuit with the important voltage waveforms. The input d-c pulses *A*, are constant amplitude pulses supplied from the first Schmitt trigger of Fig. 4 and are applied to capacitor *C*1 which charges negative through *R*1 and *CR*1. The discharge path is primarily through *R*2, which is much smaller than *R*3, and *CR*1 being a high impedance for the discharge current. This charge and discharge action produces the waveform *B* when a pulse of audio input is applied to the circuit. As *R*3 is a large resistor the charging of *C*2 produces a linear waveform as shown at *C*. At the end of the pulse of audio, *C*1 discharges rapidly through *R*2 while *C*2 discharges slowly through *R*2 and *R*3 in series. The input impedance of *Q*2 is high compared with *R*2 plus *R*3 and can be neglected. When the charge on *C*1 falls below the charge on *C*2, diode *CR*2 conducts and both *C*1 and *C*2 finish discharging rapidly through *R*2.

The firing level of the second Schmitt trigger of Fig. 4 is adjusted to provide the desired operate time delay and release time delay as shown on curve *C* of Fig. 5. The output pulse of the Schmitt trigger operates the signal gate as explained previously.

Referring again to Fig. 2, the signal detector is designed to pass only frequencies in the band of 1,700 to 2,900 cps and the noise detector is designed to reject the signal band and pass all other frequencies. The firing level of the noise detector Schmitt trigger circuit is set to operate if white noise, restricted in bandwidth to 300 to 3,500 cps, exceeds a ratio of 10 db as compared to the signal operating level of the signal detector. The time delays of the noise circuit are adjusted to operate faster and release later than the signal circuit time delays. This method of operation, while not completely foolproof, provides a high degree of voice guard immunity. Referring again to Fig. 3, even though energy distribution of many common sounds contains peaks within the 1,700- to 2,900-cps signaling band, these energy peaks must now contain 10 db more energy than the combined out-of-band frequencies.

However, there are speech sounds which will always pass through this vice guard gate. Again, referring to Fig. 3, it is seen that for intervals of 50 ms there are common word sounds whose energy peaks are completely within the signal band, especially when the energy below 300 cps is removed. This is true especially of the common sounds such as ch, th, ss, and ee. These voice sounds are effectively removed by the combination of multiple frequency tones and the second time delay circuit as shown in Fig. 1. This time delay circuit is similar to that described in Fig. 5 having a controlled operating delay and fast release. The operation of this part of the circuit requires that two frequencies which identify a digit must be present simultaneously for a fixed time duration to be registered as a digit in the following circuits.

Fig. 6 is from a photograph showing the delay times of operation of the noise guard circuit. The curves are identified as follows:

1 AGC amplifier output. Input is compound tone composed of 2,300 and 2,500 cps of equal amplitude.
2 Trigger voltage that opens the signal pass gate.
3. AND-gated outputs from the 2,300 and 2,500-cps signal detectors.
4. Trigger voltage that operates the register digit gate.

The delay and release times of the gate control circuits were adjusted on a compromise basis as the different performance requirements demand different time delays for optimum performance. These requirements break down as follows:

1. *Signal detection*—50-ms pulse minimum, requires operate delay time of 50 ms maximum, requires release time short to reset circuits for next digit as soon as possible.
2. *Voice rejection*—requires operate time delay as long as possible; a long release time delay also helps.
3. *8-ms break in signal pulse*—no effect on operate time, requires relatively long release time.

To provide optimum operation, the delay of the first gate control circuit has been adjusted to 20 ms under conditions that are normal. The rise time delay in the narrow bandpass detection filters is 8 ms and the delay of the register storage control has been adjusted to 14 ms. Under nominal conditions, therefore, the detection of a digit is accomplished in 42 ms. The release times reset the circuit

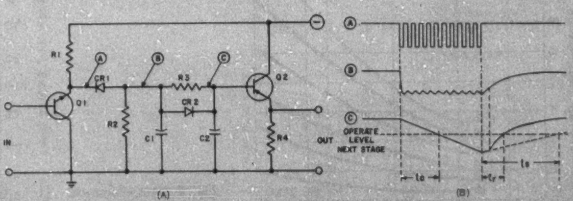


Fig. 5. A—Time delay circuit schematic. B—Voltage waveform

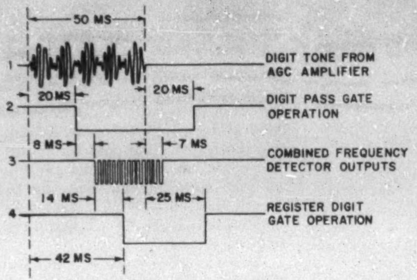

50 MS
1
DIGIT TONE FROM AGC AMPLIFIER
20 MS
20 MS
2
DIGIT PASS GATE OPERATION
8 MS
7 MS
3
COMBINED FREQUENCY DETECTOR OUTPUTS
14 MS
25 MS
4
REGISTER DIGIT GATE OPERATION
42 MS

Fig. 6. Operation of digit detector showing voice guard delays

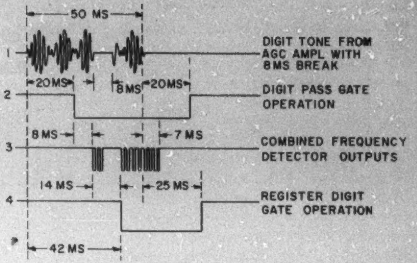

50 MS
1
DIGIT TONE FROM AGC AMPL WITH 8MS BREAK
20MS
8MS
20MS
2
DIGIT PASS GATE OPERATION
8 MS
7 MS
3
COMBINED FREQUENCY DETECTOR OUTPUTS
14 MS
25 MS
4
REGISTER DIGIT GATE OPERATION
42 MS

Fig. 7. Operation of digit detector showing 8-ms digit break

in 25 ms, thus allowing for a rapid detection of the next digit.

These delay times were measured under varying conditions of temperature, beta, and battery as required by the Signal Corps specifications. Measurements were made with varying combinations of the three parameters over the following ranges:

Temperature............................ ..20 F (degrees Fahrenheit) to 135 F
Beta.........................40 to 100
Battery....±10.0 VDC to ±14.5 VDC

Varying beta and/or battery under any fixed temperature conditions had a relatively negligible effect on the time delays in the circuit. Extreme variations of these two parameters produced a total over-all change in the detection and release time of approximately 5 ms.

Temperature, however, has a more pronounced effect on the time delays. At high temperatures the operate time of the two gates reduces to 10 and 7 ms, respectively, with the over-all operate delay time reducing from 42 ms to 25 ms. The operative time of the narrow-band-pass filters remains relatively constant, all the changes taking place in the voice guard and register storage time delay circuits. At the low extreme of temperature the operate delay time of the gates increases to 25 ms and 17 ms, respectively, allowing an over-all storage delay time of 50 ms.

Tests have also been made, under all the above environmental conditions, showing the action of the delay circuits when the input pulse is broken for 8 ms. Fig. 7 is from a photograph showing a break in the input pulse. Even though the third trace shows a break in the digit detector output, the gate control has operated only once and therefore only one digit is registered. The 8-ms break was moved throughout the 50-ms signal period. When the 8-ms break occurs within the first 20 ms of the signal, the effect is to increase the delay time for the operation of the digit storage gate. When the 8-ms break occurs during the latter 10 ms of the signal, the effect is to reduce the recovery time.

As digit reception and detection are required from telephones at various distances from the switchboard, an artificial line was constructed to simulate various lengths of telephone lines. Fig. 8 shows the attenuation characteristics of this line.

The operation of the AGC amplifier and all the time delay circuits depend upon the power in the compound digit tone, and not upon the level of any individual frequency. The frequency detectors, however, must each detect their frequency from the compound tone. As seen from Fig. 8, each detector must operate over a level range of 9 db. To accomplish this, each narrow-band-pass filter is followed by a Schmitt trigger circuit adjusted to operate over a level range of 9 db. To stay within the 9-db range produced by the attenuation characteristics of the line with frequency, it was necessary to specify that a compound tone produced by the telephone always be of such a character that the higher frequency component be either equal to or higher in level than the lower frequency component. The telephone digit sender meets these specifications and all tests were satisfactory.

Voice guard action was checked by counting digit simulations per speech hour. To obtain consistent results throughout the design and testing of the circuit, a standard voice test tape was made. This was a 30-minute tape recorded with various male and female voices. English of various accents and

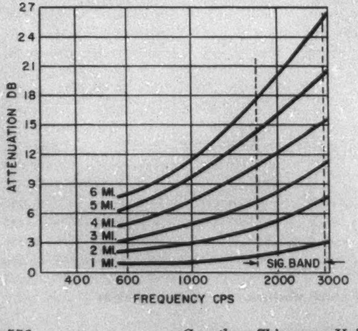


Fig. 8 (left). Attenuation characteristics of artificial line

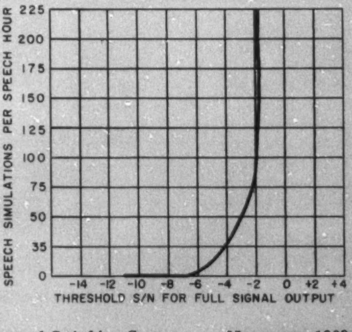


Fig. 9 (right). Speech simulations versus signal to noise ratio setting several foreign languages were used. To simulate actual operation as closely as possible, high fidelity techniques were used for recording and playback. A pulse counter connected to the digit register control gave an accurate count of the digit simulations caused by speech. Final results of voice simulation measured over the environmental range show the following results:

at 20 F = 0 digits/speech hour
76 F = 4 digits/speech hour
135 F = 16 digits/speech hour

It is interesting to note that under the same extremes of environmental conditions, the S/N of the circuit varies from 10 db to 5 db due to changes in sensitivity of the Schmitt trigger circuits. This same variation of speech simulations is observed at room temperature when the S/N of the circuit is varied. This is shown in Fig. 9.

At a sensitivity setting of 10 db or better, S/N, the circuit is immune against normal voice. As the S/N is reduced the noise gate becomes increasingly ineffective against voice, and simulations rise rapidly.

While the circuit approaches complete immunity from false operation due to voice, it should be pointed out that deliberate talkdown of the circuit is possible, although very difficult. Reference to the voice spectrograms in "Visible Speech" shows the proof of this. With practice, sustained compound sounds representing any digit may be produced by the human voice. When the input signal contains digit tones which are 10 db stronger than all extraneous tones, by definition the circuit must register a digit, whether the tones are produced by a digit sender or not.

References

1. Visible Speech (book), Potter, Kopp, Green. D. Van Nostrand Company, Inc., Princeton, N. J.

2. Considerations Pertaining to the Design of an Electronic Telephone Switching System, B. Brightman, M. P. Tubinis. *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 78, 1959 (Jan. 1960 section), pp. 777–86.

3. Transmission Aspects of an Electronic Switchboard Employing Time-Division Multiplexing, J. C. Perkins, Jr. *Ibid.*, pp. 949–53.

Production Design Considerations in Television Chassis Sweep Interference Reduction

G. W. CARSON
ASSOCIATE MEMBER AIEE

PRESENT-DAY television (TV) receivers utilize circuits that generate an appreciable amount of radio-frequency power. These signals exhibit the same characteristics as those which would be present if they came from a miniature transmitter, and they are capable of causing interference with other radio receiving devices.

Signal-generating devices are known to generate harmonic signals in addition to a signal of the desired frequency. The high-level pulse generator of the horizontal deflection circuit of TV receivers is such a device. This deflection circuit generates a fundamental pulse at 15,750 cycles per second, plus a multitude of harmonics. The harmonics are objectionable in that they extend through the amplitude-modulation broadcast band (550 to 1,600 kc) and into other much higher service frequencies.

With the advent of numerous radio and communication services throughout the radio spectrum, the problem of mutual radiation interference becomes ever more complex. Extensive research and experimental work has been conducted by committees of the Institute of Radio Engineers (IRE) in conjunction with the various industries. As a result, the IRE Committee on Radio Interference has formulated standards which define methods of measurement of spurious radiation. This work guided the FCC (Federal Communications Commission) in establishing limits and standards that the industry must endeavor to meet for efficient utilization of the already crowded radio spectrum. The present discussion of horizontal-sweep radiation reduction in TV chassis is concerned with the factors which are important to the engineer engaged in TV receiver design.

Initially, the assumption is made that the standard test setup is available; see Fig. 1. The IRE Standard[1] specifies:

1. An isolating input coupling pad matching the typical 300-ohm receiver-tuner input impedance; see Fig. 2.
2. A low-pass power-line filter network (*1805*); see Fig. 3.
3. The primary 60-cycle a-c power-supply voltage maintained at 117 volts ±2%.
4. The field strength meter adjusted so as to present a nominal 50-ohm input impedance. The function switch of this meter is set to field intensity, continuous wave, or carrier position for measurements.
5. The shielded screen enclosure at least 7 feet high, by 7 by 10 feet inside dimensions, with the test chassis a minimum of 30 inches from the walls of the shielded enclsure and 30 inches from the shielded floor structure.

The FCC and the IRE representatives from the radio and TV manufacturers

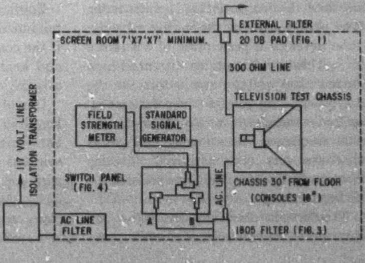


Fig. 1. Block diagram of standard setup for measurement of line-conducted interference

---

Paper 60-812, recommended by the AIEE Broadcasting Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting, Atlantic City, N. J., June 19–24, 1960. Manuscript submitted May 12, 1959; made available for printing May 9, 1960.

G. W. Carson is with the Zenith Radio Corporation, Chicago, Ill.

The author especially wishes to express his thanks to N. W. Aram, A. Cotsworth, and R. O. Gray, all of the Zenith Radio Corporation, who contributed helpful suggestions for this paper.

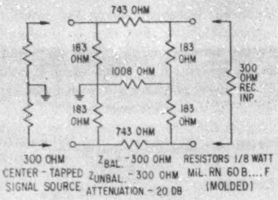


Fig. 2. Standard antenna-coupling pad of a 20-decibel attenuation and a 300-ohm impedance

are in agreement on the power-line-conducted measurement technique. In particular, it is felt that when no frequency component of the line-conducted voltage on either side of the power line is in excess of 100 microvolts, the direct field radiation to other TV and radio services is not likely to cause objectionable interference. This is contrasted with previous procedures which required free-field loop measurements in isolated interference-free locations remote from the city.

Extensive research and measurements among various TV receiver manufacturers have reduced the deviations between test setups, made according to Fig. 1 to a point of acceptable agreement. Correctly designed screen rooms and equipment are mandatory for obtaining reproducible data. The Stoddart Aircraft Radio Company, Inc., *NM20B* radio-interference and field intensity meter has been found satisfactory for these measurements, although other units of comparable quality are available.

A coaxial switching assembly has been very useful in maintaining the terminal impedance of the unused measuring points; see Fig. 4. A standard signal generator should be available as shown in Fig. 1, to cross-check the accuracy of measurements at regular intervals. This is necessary because of the idiosyncrasies of even the better field intensity meters.

Analyzing the TV chassis reveals two prime causes of horizontal sweep harmonic interference: the electrostatic field of voltage and the magnetic field of currents. The plate circuit of the horizontal pulse tube develops a peak pulse of 3.5 to 5.5 kv. This circuit contains extremely high current and voltage components at the 15,750-cycle horizontal-sweep frequency, and substantial amounts at harmonics of this frequency. Random-dressed wires that are very close physically to this circuit will therefore exhibit excessive interference levels.

Troublesome chassis designs in the past have permitted the chassis base to perform as an antenna. A measurable loop voltage in these cases existed between the chassis and the screen room or adjacent grounded system. The amount of this effect is influenced by the antenna input coupling impedance, the chassis and circuit capacitances, and the inductive coupling to the a-c power cord. Two basic design configurations connect the line power to the tube heaters and the rectifier *B* supply. In the one, transformers couple the proper voltages and currents to the tube heaters and rectifier circuit; in the other, the heater filaments are connected in a series sequence directly across the power line, as are the selenium or silicon rectifiers.

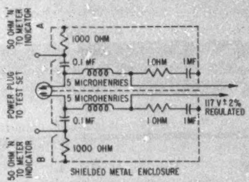


Fig. 3. Standard power line impedance filter (1805)

By virtue of the inherent isolation of the power transformer secondaries from the line in the first case, the capacitive and inductive harmonic interference is held to a minimum. Series-filament chassis with selenium or silicon rectifier power supplies present a difficult problem because of the direct electric connection to the power line cord. With this design, it is imperative that low-harmonic voltage appear between the chassis and ground. Basic circuit design and chassis layout must be well chosen to minimize these undesired harmonic voltages on the chassis. When cabinet design provides a metallic enclosure or cabinet, this directly reduces the objectionable harmonics. Wooden cabinets or any of the synthetic substitutes do not shield the sweep harmonics. An attenuation of several decibels is attained because of the shielding effect of a metal cabinet.

Spurious sweep radiation can best be minimized in original TV receiver design if the contributing factors are recognized and kept in mind during layout of the original circuit.

One of the prime evils is excessive lead length in sweep circuits. Leads carrying harmonic sweep frequencies will couple spurious frequencies to other chassis wiring if located within fractions of an inch. A most critical lead is the top plate connection to the horizontal-sweep tube. This tube should be placed as close as possible to the sweep transformer, which is normally in a shielded enclosure. The shield cage should be well grounded electrically whether screws or twist lugs are used. Inadequate grounding of the shielded sweep cage gives rise to large harmonic sweep interference currents at the high-impedance mounting point where it is attached to the chassis.

The sweep circuit (tubes and transformer) should be isolated physically from the rest of the chassis components to minimize coupling to other circuits and to avoid common ground coupling as well as direct electrostatic coupling.

Circulating ground currents are best minimized by short ground leads. Avoid the common ground connection from a cathode terminal to a filament terminal and then to ground. This introduces the peak saw-tooth current from the horizontal tube into the primary filament wiring by virtue of common impedance. This, in turn, couples and radiates to other wiring and eventually the interference is conducted to the power line.

Surprising reductions are effected by attention to the total physical area of the sweep circuitry and to the wire lead lengths. Parts should be located adjacent to the circuit concerned. The total area of the entire circuit should be as small as possible. In addition to reduced radiation from short leads, circulating ground currents are thereby reduced. This minimizes common coupling to nearby components.

The electrolytic capacitor of the power supply filter is very important in the series filament selenium-rectifier receiver design. This component should maintain low impedance to the saw-tooth current flowing in the horizontal-sweep tube circuitry. The multiple-section electrolytic capacitor should be checked for low-coupled impedance at the sweep harmonic frequencies in the broadcast band. The filter capacitor should be physically close to the sweep circuits to maintain short lead lengths. These leads radiate appre-

Fig. 4. Coaxial constant-impedance switching assembly

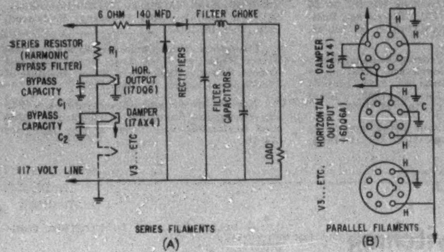


Fig. 5. **Preferred filament-wiring configurations**

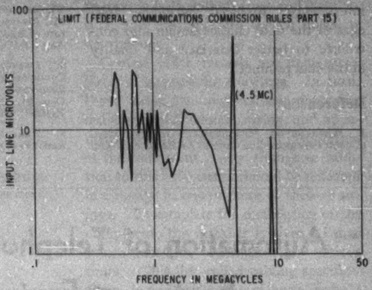


Fig. 7. **Conducted power-line emission of sweep harmonics of receiver complying with FCC requirements**

ciable harmonics and couple undesired signals directly into unrelated circuits.

Assuming that nominal effort is made to keep the wiring short and dressed away from other circuits, the deflection yoke and its connecting wires have not contributed appreciably toward excessive sweep harmonics.

Transformerless series-filament receiver design presents one problem because of the heater-cathode capacitance of the damper control tube. The sweep voltages and harmonics are directly coupled to the a-c power line through this capacitance via the series-filament wiring. Appreciable reductions in sweep harmonic line conductance are effected by the arrangement of the tube filament sequence wiring. Fig. 5 is the preferred arrangement of, first, a series dropping line resistance *R*1; second, the horizontal output tube filament; and third the damper filament.

The added series impedance gained by this arrangement reduces the harmonic voltages coupled to the filament from the damper cathode because of the heater-cathode capacity and the pulse existing in the cathode circuit.

Additional bypass capacitance (*C*1 and *C*2) from the filament circuit to ground was required in one such case to bring the sweep harmonics within the required measurement. It may appear that these horizontal sweep filaments could be placed at the bottom end of the filament string. This is ruled out because of other problems encountered if the low-level or high-frequency amplifying tubes (radio-frequency tuner or intermediate frequency) do not have the filament adjacent to the cold or "grounded" side of the series-filament 117-volt string.

It is preferable in any design to use only the one required connecting wire to each of the filament solder terminals at the damper tube socket. Likewise, only one lead should connect for the respective *B*+ circuit, as shown for the parallel-filament case in Fig. 5(B). If tube socket terminals are used indiscriminately as convenient tie points, appreciably more harmonic currents will be introduced into related circuit wiring.

The ground lead from the filament transformer should be short and away from horizontal circuits. Also, the primary a-c wires should be isolated from the horizontal circuits.

Since the Underwriters Laboratories, Inc., have increased the limit on total maximum high-voltage circuit capacitance to 3,000 micromicrofarads, this can be utilized to great advantage by increasing the bulb-coating capacitance. The rectified voltage from the horizontal-pulse transformer flows in this capacitance. Thus, more capacitance and better grounding aid substantially in reducing the high-level harmonic voltages.

To take advantage of this greater allowed capacitance, the outside graphite bulb coating was extended to the outer rim of the picture tube. The graphite coating, was grounded by the mounting strap used to secure the tube to the chassis, giving a low-resistance d-c connection to the chassis ground. Equally important is a low-impedance ground path offered by a wide metal strap to the alternating voltages in the circuit. This improvement in reducing the grounding resistance and impedance of the picture-tube bulb coating has been accomplished in Zenith television receivers with a reduction in cost; see Fig. 6. Fig. 7 shows the measured conducted power line emission.

It is axiomatic that there exist no one or two simple corrections that will remove the problem of undesired sweep harmonic interference. The solution is primarily good basic initial design in all the related parts, with the consistent aim of minimum harmonic-sweep interference.

Minor circuit modifications can increase the measured harmonics at one frequency due to in-phase addition of these voltages. It frequently happens that one troublesome circuit must be modified to accomplish, say, a 6-decibel reduction in the harmonic measurements before other relatively minor changes will give the last decibel or so needed to attain the desired low level of harmonic

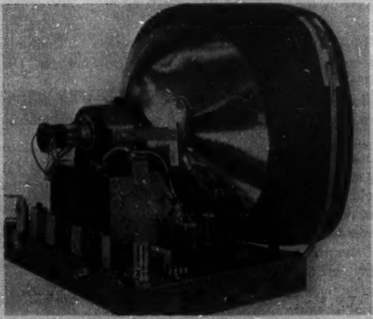


Fig. 6. **Zenith TV chassis illustrating ground-strap connection to external graphite electrode of picture tube**

interference. Reduced harmonic voltages in the final chassis design also contribute to better interlace and quality in the final product.

References

1. Methods of Measurement of the Conducted Interference Output of Broadcast and TV Receivers in the Range of 300 KC to 25 MC (Standard 56 IRE 27S1). *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 44, Aug. 1956, pp. 1040–43.

2. *Amendment*, Part 15, Federal Communications Commissions' Rules Governing Restricted Radiation Devices. *Docket no. 9288*, adopted Dec. 21, 1955, released Dec. 23, 1955; *Federal Register*, Washington, D. C., Dec. 29, 1955.

3. Measurement of Radio Interference in Accordance with Specification MIL-1-6181B, Using Stoddard NM20A or NM20B Equipment. Stoddard Aircraft Radio Company, Hollywood, Calif., Feb. 15, 1956.

4. *Proceedings*, Third Conference on Radio Interference Reduction, U. S. Army Signal Labs and Armour Research Foundation, Illinois Institute of Technology, Chicago, Ill., Feb. 1957.

5. Control of Spurious Radiation from FM and Television Receivers, R. J. Farber. *Electrical Engineering*, vol. 76, no. 5, May 1957, pp. 414–16.

Automation of Telephone Switching Equipment Engineering

L. W. BLUMER
NONMEMBER AIEE

G. E. REIFENSTUHL
NONMEMBER AIEE

N. E. LIST
NONMEMBER AIEE

THE TELEPHONE industry is beginning to use computers to engineer new telephone switching equipment and additions to it. Most of this equipment which sets up calls among subscribers is located in central offices. An examination of the factors involved in manually engineering this equipment will help to explain why Bell Telephone Company engineers feel computers will make engineering faster, cheaper, and more accurate.

What Is Involved in Engineering of Switching Equipment?

Engineers must first determine which type of central office switching equipment is to be installed. Office size, trunking to other offices, service objectives, maintenance costs, expected growth rate, economics, and many other factors bear on this decision.

Assume for illustration that crossbar dial equipment is chosen to replace an existing manual office. The number of subscribers at "cutover" time must be estimated. Calling rate and call duration time must be computed on the basis of locally collected data and the projections made from them. Consideration of these factors, capacities of various equipment required and service objectives, etc., fix the minimum size of the switching system.

Economics of providing building space and equipment for growth help fix the maximum office size provided initially.

In order to price out the job, the engineer must know engineering, equipment, and installation costs. He must determine the time intervals required for each to establish a schedule for the job.

The engineer must consider the variables covered above and many others which interrelate with them. He must make a series of decisions based on these factors. For large dial jobs he usually writes an outline specification. He makes cost estimates and obtains management authorization for the undertaking.

For the job just described the Bell Telephone Company engineer then places an order with the Western Electric Company to detail engineer, manufacture, and install the equipment.

His outline specification, order, and job drawings specify:

1. Where, when and what equipment is to be installed.
2. Floor plan layout.
3. Trunking patterns, assignment of trunks to switching equipment.
4. Equipment features.
5. Anticipated growth patterns.

The Western Electric engineer's subsequent job in preparing detailed specifications and job drawings is covered later.

Why Consider Computers?

"Manual" engineering takes considerable time. Engineering rates loaded for supervision, clerical and drafting assistance, floor space, etc. have risen over the years.

Time and costs involved often preclude consideration of all desirable alternatives. Computers are fast. They could permit checking reasonable alternatives quickly.

Many elements, such as conversion of calling rate and call duration data into equipment quantities, are highly repetitive computations. Computers are designed for this work. Engineering talent could better be used to program computers.

Humans thrive on innovations. Engineers excel. But this sometimes adversely affects job uniformity. Computers do not innovate.

Faced with the same choice of alternatives at different times, humans are not consistent. Computers are.

Engineering rules must be learned and remembered. Although they must be taught in very simple steps, computers learn, remember, and apply what they know much better than humans. So they offer the advantages of accuracy and consistency to engineering.

Summed up: Because of experience with computers and knowledge of their capabilities, Bell engineers feel they will offer cost reduction, interval reduction, standardization, uniformity, and other advantages to equipment engineering.

Will Computers Be Feasible for Engineering?

Yes. Bell Telephone Company engineers have this evidence:

1. Rapidly expanding use of computers to write detailed central office equipment engineering specifications at Western Electric. About 15% of all job specifications written there are now produced by computer. Details are given in the section on Automation of Western Electric Equipment Engineering.

2. Computer preparation of detailed specifications for large business customer dial switching systems. These data will be discussed in the section on Computer Processed Short Order Plan for Dial PBX's.

3. Telephone company feasibility studies. Gross savings in engineering

Vol.
Pt. 1
1960

Paper 60-817, recommended by the AIEE Communication Switching Systems and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting, Atlantic City, N. J., June 19–24, 1960. Manuscript submitted March 21, 1960; made available for printing April 27, 1960.

L. W. Blumer is with the Northwestern Bell Telephone Company, Omaha, Neb. G. E. Reifenstuhl is with the Western Electric Company, Chicago, Ill. N. E. List is with the Michigan Bell Telephone Company, Detroit, Mich.

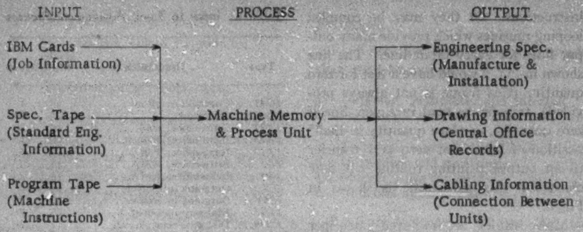


Fig. 1. Line-by-line specification writing

man-hours up to 60% are expected. Further details will be found in the section on Feasibility Studies of Computers Used for Engineering.

4. Limited use of punched card techniques of engineering, which will be discussed later in the paper.

Automation of Western Electric Equipment Engineering

Bell Telephone Companies prepare outline specifications on central office equipment which is to be detail engineered, furnished, and installed by the Western Electric Company. This section discusses some of the techniques which have been developed for the production of detailed engineering information by a large scale electronic computer used for many other purposes.

The Western Electric equipment engineer performs three basic functions in his detailed analysis of the telephone company specifications. The automated procedure must provide information for these functions in printed form for further manual processing or in the form of cards or tape for further machine processing. These functions are as follows:

Information for manufacture
Information for installation
Maintenance of central office records

These functions are performed by engineers at two locations on both manual and dial switching equipments. Toll dial switching equipment is engineered in Newark, N. J. Regular central office dial switching equipment is engineered in Chicago, Ill. At the present time much of the engineering in Chicago is concerned with the system known as No. 5 crossbar. This system makes use of crossbar switches and relay common control circuitry to provide talking paths among subscribers. The techniques described in this section have been generally applied to this system. Some attempts have been made toward applying them to step-by-step, No. 1 crossbar, and crossbar tandem dial systems. However, these systems are older. They are not as adaptable to engineering automation techniques as the more standardized No. 5 crossbar system.

In the manual engineering of a No. 5 crossbar order, manufacturing information is provided by a series of specifications. In general these specifications cover a type of frame or unit, and contain a summary of materials required. The major equipments are ordered as lists on manufacturing drawings. These lists have been provided for the various standard operations which the equipment has been arranged to perform. Combinations of types and quantities of equipment provide central office installations capable of fulfilling almost any traffic requirement. Another portion of these specifications includes a listing of the drawings to be used in both manufacture and installation, and contains any special notes or instructions.

Central office records are kept on a series of engineering drawings. In turn, these show the various manufacturing drawings, with figures, wiring, and apparatus used in production and installation of the equipment. Any change or addition to this equipment should be reflected in a change in one or more of these drawings. To complete the installation of the equipment, the installer is given a list of interconnections between the various units. This list, called cabling specifications, includes number of runs, switchboard cable code, length of run, and termination on the units.

It can be seen from the foregoing description that certain information will appear in a number of forms throughout the order. So certain basic decisions will logically lead to a need for detail information in many places. As an example of this statement, assume two frames (0 and 1) are being added, with an option A on Frame 0 and an option B on Frame 1. Since these frames will be ordered by a series of manufacturing drawings, it can be determined what these drawing numbers should be and may add or delete certain drawings on the basis of option A or B. The summary of materials will require lists for the manufacture of the basic frames and Frame 0 will require a list for option A while Frame 1 will require a

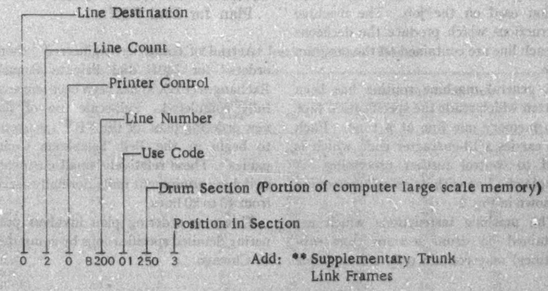


Fig. 2. Line coding

The codes are used as follows:
Line Destination—Number of cards required for this line
Line Count—Number of lines being printed. Provides count for page and title routine
Printer Control—Controls operation of the printer. A zero provides double space, a one a new page, etc.
Line Number—Provides a means of sequence checking
Use Code—Used to control further processing on the basis of whether this line is always required, or needs modification, etc.
Drum Section—If line required modification, machine instructions will be found within this drum section
Position in Section—The instructions start at this tenth of the 200-character section

different list for option B. These options will be reflected in the central office record drawings. The options may appear as differences in cabling between these frames and other units. Production of detailed information from relatively few basic engineering decisions makes computer production of these details practical.

Early investigation indicated that even a large scale computer would not have the memory capacity required to write a complete specification. This led to a line-by-line approach to specification writing. This concept is shown in Fig. 1.

The first step in preparing a specification for production by this method is an analysis of the various equipment engineering decisions which may affect the specification. After this analysis has been completed a master specification containing lines for all anticipated conditions is prepared. This will later be placed on tape and becomes the input specification tape of Fig. 1. During the writing of the master specification, notes are made of the engineering decisions which must be made for each job. These notes are made into a questionnaire which will be used to produce the job input punched cards. The lines within the master specifications can be divided into three general types: 1. lines which are always required and will not be modified by the job; 2. lines which are always required but will require modification per job information; 3. lines which may or may not be required and may require modification based on the option used on the job. The machine instructions which produce the decisions on each line are contained on the program tape.

A general machine routine has been written which reads the specification tape into memory one line at a time. Each line carries a 13-character code which is used to control further processing. A sample of a line with its associated coding is shown in Fig. 2.

The machine instructions which are contained on drum sections (line subroutines) may contain only a few simple instructions. Or they may be complex looping routines which provide many output lines for every input line. The line shown in Fig. 2 would have a test for zero quantity (this frame is not always provided), transfer to read in a new line if zero quantity, or place quantity in the** position of line if not zero and transfer to an output printing routine. A flow chart for this subroutine is shown in Fig. 3.

Many cables are required once per frame type. The line routine for such a condition would be transferred to the general printing routine, but would then be transferred back to the same line instead of bringing in a new one. This line would be modified to raise the frame number, and then printed again. A count would control the number of times this was done, and a new line would be brought in when all cables had been written.

Every line would be read and processed against the order cards for the specification. Upon completion of the specification, a new set of order cards would be read into memory. The lines for this specification type would be found and processed one at a time. When all order cards have been processed, the general routine will rewind the tapes and type out a message for the machine operator.

Describing the process requires considerable time. However, an average specification can be machine processed in a minute to a minute and a half.

Computer Processed Short Order Plan for Dial PBX's

A trial of computer-engineered "short orders" for 740E dial Private Branch Exchanges (PBX's) has been been successfully completed. Full-scale use of the new ordering plan for the PBX's is about to begin in the Bell Telephone Companies. These relatively small customer switching equipment units normally serve from 40 to 80 lines.

This new ordering plan involves preparing detailed specifications by computer in Chicago.

Table I. Input to Trunk Assignment Process

Type	Description	Quantity*	Group
1G43	Common overflow	26	1
Spec	See below	26	2
1A11	Intra-office; coin	10	3
1A13	Intra-office; F.R.	60	4
1A13	Intra-office; F.R.	60	5
1A1	Intra-office; AMA	14	4
1A1	Intra-office; AMA	16	5
1C6	Outgoing interoffice	96	6
1C6V	Outgoing interoffice	32	7
1D15	Dial zero non-coin	26	7
1D56	Dial zero coin	12	7
1G45	Combination tone non-coin	10	7
1C166	Coin junctor	10	7
1C16k	Interoffice coin	8	7
1G51	Permanent signal hold	8	7
1G44	Combination tone; coin	6	7
1D100	Time announcement	6	7
1D106	Information	6	7
1D56D	Vacant code intercept	6	7
1D94	Emergency announcement	6	7
B61M	Incoming; MF pulsing	106	8
B90D	By-link non-tan. Inc.	20	8
B115M	Toll switch; coin; cont. ring	14	8
B61U	Interoffice incoming	10	8
	Special trunks consist of:		
B112C	Toll switch no test	2	
Bunch	Appearances for coin junctor	2	
D122	Repair: day; trans. cont.	1	
1D109	Repair: night trans.	2	
1E16	Outgoing to LTD	4	
1E24	Incoming from LTD—N.T.	4	
1E41	Ringer test	4	
MT	Master test frame	1	
1A18	Reverting code	6	
		26	

* Distribution made over 5 frames.

An order originates in a Bell Telephone Company with preparation of a 3-page form and requisition sheet. The engineer simply fills in blanks or encircles options. Thus, he orders in outline form, not in technical description, the number and types of switches, the switchboard, etc.

As many copies as required are made by any quick reproduction method available in the telephone company. Copies are used as needed locally. Others are sent to appropriate distribution, installation, and manufacturing organizations of the Western Electric Company.

At Chicago, a deck of punched cards is prepared from the quick order form. These cards, master specification tape, and a program tape are loaded into the computer. The output is a complete detailed specification which combines the following:

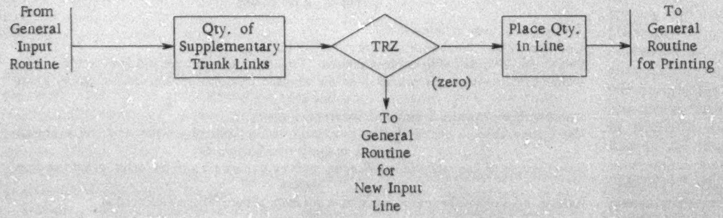

From General Input Routine
Qty. of Supplementary Trunk Links
TRZ
(zero)
Place Qty. in Line
To General Routine for Printing
To General Routine for New Input Line

Fig. 3. Subroutine

1. The general job summary.
2. Material list, edited by sources.
3. Cabling specification.
4. Job drawings in tabular form.

Copies of the completed specification are used by the manufacturing organization in filling the order. Others are used by the installer and the distribution people for billing and other uses. Of course, copies go to the telephone company.

The advantages of this quick order plan are:

1. Shortening of the order-to-service interval by reducing the telephone company engineering time. The short order may be filled out in minutes. Writing the detailed specification manually takes days and is subject to human errors that occur even in the most careful detail engineering.

2. Engineering costs will decrease attractively.

3. Standardization offered by computer processing will provide installation, manufacturing, billing, and other difficult-to-evaluate but real values.

4. The engineer is "out of the nuts and bolts business." This releases his energies for more productive efforts requiring his engineering talents.

5. The ordering form serves as a valuable check list since every question on it must be answered. So delays in jobs due to omissions in ordering should decrease materially,

Feasibility Studies of Computers Used for Engineering

As part of a test of computer practicability, one telephone company made a feasibility study on computer assignment of trunks to trunk link frames in No. 5 crossbar offices.

The studies were directed by N. E. List and developed by G. L. Broxholm, both with the Michigan Bell Telephone Company. This phase of the outline specification is considered among the most complex and time consuming jobs in preparing an outline specification. The assignment routine was programmed for a punched card input–output type computer. The following paragraphs describe the nature of the problem, input requirements, and a reproduction of the machine listing of the output cards produced by the computer process.

OBJECTIVE

The objective is to distribute all trunks as evenly as possible over all frames to insure that as many calls as possible can be completed without blockages.

PROBLEMS

1. There are several types of trunks that must be assigned:

a. Outgoing, interoffice, and intertoll.
b. Incoming, interoffice, intertoll, and tandem.
c. Two-way, intertoll, and CDO.
d. Intra-office and reverting.
e. Marker pulse conversion.
f. Bunched appearances.

2. For equipment reasons the 160 possible appearances on a trunk link frame are divided into 16 levels of 10 appearances each. Each level has certain limitations:

a. 8 levels can be used only for outgoing trunks and the outgoing end of intra-office trunks.
b. 4 levels can be used only for incoming trunks and the incoming end of intra-office trunks.
c. 4 levels can be used for any type of trunk except intra-office.
d. The 12 levels (items a and c) that can be used for outgoing trunks are subdivided into six groups of 2 levels each. With some exceptions, once any trunk of a type is assigned to one of these six groups all the trunks of this type must be assigned to the same group.

INPUT INFORMATION

1. Number of trunk link frames.

2. Number of vacant terminations on each level of each frame and totals for each level and frame.

3. Order of preference of assignment to levels for each group of trunk types.

4. Trunks to be assigned by type, quantity, and group number; see Table I.

The machine process distribution of trunks shown in Table II is felt superior to that normally produced by manual methods.

Projection of these and other feasibility study results to the whole telephone company engineering job indicates a 60% engineering manhour saving over manual methods.

Punched Card Engineering

Detailed job specifications for small new step-by-step dial offices and additions to such offices have generally been prepared by telephone company engineers.

Engineers in at least two Bell Telephone Companies have developed techniques to use punched card processing business machines for detail engineering of small dial offices.

Basically the procedures combine elements of those covered in the foregoing sections. The engineer prepares work sheet forms. He enters the codes, quantities, and associated information pertaining to the desired equipment items.

The machine process produces:

1. The installer's cable running list.
2. Summary of material required.
3. Equipment costs.
4. Installation man-hour estimates.
5. Other elements of a complete specification.

Table II. Output of Trunk Assignment Process

Trunk Type	LV	TB	Quantity Trunks	Trunks Per Trunk Link				
				00	*01*	*02*	*03*	*04*
1G43	A2	0	26	04	06	04	06	06
1A13	A6	4	50	10	10	10	10	10
1A13	A7	4	10	02	02	02	02	02
1A13	A8	5	50	10	10	10	10	10
1A13	A9	5	10	02	02	02	02	02
1A1	A7	4	14	04	02	04	02	02
1A1	A9	5	16	02	04	02	04	04
1A11	A9	5	10	02	02	02	02	02
Spec.	B5	3	26	06	04	06	04	06
1C6	A3	1	50	10	10	10	10	10
1C6	A4	2	46	10	08	10	08	10
1C6V	A5	3	32	06	08	06	06	06
1D15	A7	4	26	04	06	04	06	06
1D56	A5	3	12	04	02	02	02	02
1G45	A9	5	10	02	02	02	02	02
1C166	B3	1	10	02	02	02	02	02
1C161	B3	1	8	02	00	02	02	02
1G51	B3	1	8	02	02	02	02	00
1G44	A5	3	6	00	00	02	02	02
1D100	B3	1	6	02	02	00	00	02
1D106	B3	1	6	00	02	02	02	00
1D56D	B3	1	6	00	02	00	02	02
1D94	B3	1	6	02	00	02	00	02
B61M	B9	*	14	04	02	04	02	02
B61M	B7	*	26	04	06	04	06	06
B61M	B2	0	50	10	10	10	10	10
B61M	B5	3	16	04	04	02	04	02
B90D	B5	3	8	00	02	02	02	02
B90D	B4	2	12	04	02	02	02	02
B115M	B4	2	14	02	04	04	02	02
B61U	B4	2	10	02	02	02	02	02

Steps in the Job

Bell engineers feel that much further documentation of engineering information is necessary for most effective automation. Many engineering rules previously used must be closely evaluated and defined before computers can apply them. Engineering procedures will be integrated with related clerical, accounting, and other functions.

Engineering will be a profitable part-time computer job. So these machines will largely be proven-in to handle other, more large scale data processing jobs. Hence, the engineer will not usually have direct choice of the computer available to him. But the kinds and types of computers used by the various Bell Telephone Companies will be limited.

Study of the problems involved in mechanization of engineering has suggested joint intercompany development of engineering programs for computers. This is already underway. The output of such joint effort will be modified as required by the using company to fit its computer and its organization.

Conclusions

The computer will free telephone company engineers from many tedious quasi-engineering tasks. Costs of engineering should be reduced or held in check. Overall intervals will decrease. Accuracy will improve. Resulting standardization will be responsible for further savings. By-products possible—such as cost data, accounting information, power drain data, etc.—may be even more valuable than the tangible advantages seen today.

A Step-Driven Digital Magnetic Recorder

J. G. HANSEN
NONMEMBER AIEE

THE DESIGN and installation of a freight-car information sorting system for the Chicago, Rock Island and Pacific Railroad was recently completed by the Stromberg-Carlson Company. The system, located in the Chicago, Ill., offices of the Rock Island Line, receives information on train make-up from several distant marshalling yards via teletypewriter links, records these messages, and sorts them. The Wheel Report Sorting System, as it is called, sorts the incoming information according to state of origin of shipment and state of destination, and routes this information to teletypewriters assigned to these states. In the course of designing the sorting system, a requirement developed for a means of storing multiple incoming messages efficiently, while sorting units are busy. The storage means must permit simultaneous recording and read-out at different rates, and must also operate intermittently on read-out due to switching of output during sorting. This service can be provided by paper-tape punches and readers with a buffer loop between. However, the information density on paper tape is low, thus creating a need for a physically large buffer storage. This fact, coupled with the need for continuously replenishing the tape, and the maintenance requirements of punches, led to the search for a more suitable means of storage. Magnetic recording presents the possibility of higher information density in addition to using a reusable medium. However, magnetic tape recorders of the conventional capstan-drive type do not generally provide for differing speeds of simultaneous recording and read-out with a buffer loop between for "first in, first out" storage. In addition, the requirement that the read-out be intermittent due to the setting up of connections to different printers is difficult to achieve without sacrificing information density for starting and braking space.

Description

After reviewing the market and not finding a suitable recorder available, the company's Research and Advanced Development Division proceeded to develop a magnetic recorder specifically for the Rock Island Wheel Report Sorting System. A magnetic recorder which would incrementally move the recording medium for each character being recorded or read out, on demand only, offers a means of obtaining varying information rates and instantaneous starting and stopping as required by the system. Storage between recording and read-out could be obtained by the use of a random-fold bin to accumulate the amount of recording medium required.

In order to obtain incremental, or stepping, movement of the recording medium, a simple and rugged mechanism is most desirable. The ratchet and pawl is such a mechanism and was chosen as the driving means for this recorder. There are two driving mechanisms used on each recorder; one for recording and the second for read-out. Construction features include a heavy base casting upon the rear of which are mounted the driving magnet and pawl assembly, and the ratchet gear. On the face of the castings are mounted the recording head assembly, film guides, and 35-millimeter-film drive sprocket. The ratchet used is a 160-tooth gear. This ratchet is directly coupled to the 16-tooth 35-millimeter-film drive sprocket giving ten characters of recording between sprocket holes in standard 35-millimeter magnetic recording film. Longitudinal bit spacing is held to 0.01875-inch centers (approximately 53 bits per inch, per track). Wearing surfaces of the ratchet teeth and the pawl are hardened to minimize wear. The pawl is made "glass-hard" and the ratchet teeth somewhat softer to obtain a match of wearing qualities which extends the life of the combination beyond that which would occur with both teeth and pawl made "glass-hard." Accelerated life tests on the recorder drive mechanism indicate life expectancy of at least 20 million operations, equivalent to continuous operation for approximately 750 hours in teletypewriter service at 75 words per minute, before requiring overhaul. The base castings are of cast iron to provide rigidity for mounting heads and mechanism in accurate alignment. Mass of the castings is large (approximately 2 pounds) and is arranged to prevent the action of the magnet and armature from transmitting unwanted inertial motion to the film. Such undesirable action can cause buckling of the film under the head with possible errors in reading or recording occurring as a result.

Sprocketed film is used as the record-

Paper 60-183, recommended by the AIEE Telegraph Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., January 31–February 5, 1960. Manuscript submitted August 7, 1959; made available for printing February 17, 1960.

J. G. Hansen is with the Stromberg-Carlson Company, a Division of General Dynamics Corporation, Rochester, N. Y.

Fig. 1. Step-driven digital magnetic recorder

ing medium to insure accurate longitudinal spacing of recorded bits. Sprocket holes are held to ±0.001-inch center-line spacing tolerance in conformance with the standards of the American Society of Motion Picture and Television Engineers. The thermal coefficient of expansion of acetate film is negligible in this service as is its expansion or contraction due to moisture content. Fireproof film is, of course, used. Use of standard motion-picture film base with magnetic-oxide coating permits the use of standard motion-picture projector sprockets and guides which are also manufactured to close tolerances set by the Society.

Accuracy of bit spacing in the longitudinal direction is required in order that the relative position of each recorded bit be identical with regard to recording and read-out heads. This requirement is brought about by the fact that maximum output signal is obtained by the complete passage of a magnetized portion of film under the read head during each incremental motion of the film. In addition to yielding maximum signal, this requirement prevents false signals being generated by the leading and/or trailing edges of a recorded bit. Five-bit teletypewriter code is used in the Wheel Report Sorting System. A character is represented by a "mark" or "space" in each of five parallel tracks.

At this point it should be made clear that pulse information is recorded in each track of the film as the film is stationary under the record heads and is read off the film as it is moved past the read heads. It is important, when initially adjusting the position of the read head assembly, that the head gaps are set exactly halfway between recorded bits. The procedure used to set the position of the read heads is as follows: Recordings are made in all of the multiple tracks sequentially for several inches of film, using the recording heads. The film is then moved to the reading position where it is arranged to record again in all of the multiple tracks for a short distance along the film. The piece of film is then removed from the recorder, placed momentarily in a suspension of carbonyl iron powder in alcohol, and carefully dried. The relative positions of recording and reading heads is then visible by virtue of the concentration of iron powder at the magnetized spots. The procedure is repeated until the spots made by the read heads are midway between those made by the record heads. In order to insure identical longitudinal spacing for each of the multiple tracks, the recording and reading head groups are rigidly held perpendicular to the film travel and thereby parallel to each other.

The amplitude of pulses generated by the passage of recorded bits under the read heads is kept uniform by the mechanical arrangement of the step-drive unit. Motion of the film is accomplished on the spring return of the magnet armature to its reset position, not on energizing the magnet. This prevents the possible variation in actuating speed which can occur with variable voltage applied to the magnet coil. The action of the pull-up of the armature has then no relation to the movement of film under the heads except that frequency of action determines the average speed of film travel.

Provision for storage of messages between recording and read-out is provided by means of a bin into which the film is inserted and allowed to fold randomly after recording. Fig. 1 shows the storage bin attached to the recorder. Film is withdrawn as messages are read and then fed back into the same bin. As further messages are recorded, the film is again withdrawn by the recording mechanism, passing the previously recorded film under an erasing head, just prior to recording the new information. Thus a loop of magnetic recording film is utilized. In the case of the Rock Island system, 60 feet of film are used in each loop, enough for $1\frac{3}{4}$ hours of recording time at 60 words per minute. In the system used in the report sorter, a means for detecting a taut film condition between recording and read-out is installed. This is in the form of a light metal arm, on the end of which is a roller, riding on the film as a taut condition is approached. When the film raises the arm to a preset position, read-out is suspended when simultaneous recording and read-out are proceeding, or, when read-out only is underway, the action causes the recording mechanism to step in synchronism with the read-out until the end of a message is detected.

Erasure of previously recorded information is accomplished simply by use of a permanent magnet, which biases the magnetic oxide in a condition opposite to that created by recording. This gives maximum change of condition between blank and recorded spaces yielding maximum output pulse amplitude as read out.

As the recorder is used in the Rock Island report sorting system, it operates at a speed of approximately seven steps per second. The design of the recorder is such, however, that it can be operated at 40 steps per second continuously with no modifications. Using slight modification of armature construction and the return spring, the basic unit has been driven, in tests, at 80 steps per second. The recorder as initially designed, then, can be used in service requiring the equivalent of approximately 400 words per minute of teletypewriter data. Driving energy required by the recorder is minimal, with energy being supplied to the mechanism in pulse form, at a rate determined by the data being handled. The driving magnets are energized by transistors in the sorting system from a 48-volt d-c source. The lateral spacing of the heads is not the most efficient based on the width of the film available, since only five tracks are used. With other commercially available heads, 8-track recording could be arranged across the film. If the space available outside the holes were utilized, ten tracks could be used.

Conclusions

The step-driven digital magnetic recorder is a versatile data-storage facility. It permits the simultaneous recording and transmission of digital data at varying rates. Data may be transmitted at any rate from a minimum of less than one character per minute to a maximum of 40 characters per second with no change in pulse amplitude, thus simplifying the design of amplifiers in the transmitting system. Storage of data between receipt and transmission can be as large as required by any particular system, dependent only on the size of bin which can be mechanically fitted to the circumstances. A possible additional application of the recorder is buffer storage of periodic readings of remote data sources between transmissions to a central point.

Reflections on the Traffic Usage Recorder with Emphasis on Traffic Engineering

R. A. BAUERFELD
NONMEMBER AIEE

AS IMPLIED by the title, the intent of this article is to relate the manner of use and value of the Traffic Usage Recorder as applied to telephone traffic engineering. To develop a full appreciation of this, a general knowledge of the responsibility of the traffic engineer is required. In a broad sense, the traffic engineer must see that facilities of the right kind and of the right amounts are in the right place at the right time to give the right grade of telephone service. An important phase in this provision of telephone facilities is in determining the requirements of the telephone subscriber. In the process of estimating subscriber requirements, summarized call carrying and associated equipment usage data are analyzed, projected, and modified to reflect expected growth, new service, and proposed equipment arrangements. The modified basic data are then converted to equipment quantities and trunks through the use of probability tables reflecting the desired grades of service. The first step of this process, the gathering and summarization of usage data, is of prime importance and has been a major problem to the traffic engineer.

Prior to the development of the Traffic Usage Recorder, sufficient accurate usage data was most difficult to obtain. Through the use of peg count registers, data were available that indicated the number of times an item or groups of equipment were called into service. However, this is not usage; the length of time of equipment in use was needed. To convert peg count data to usage it was necessary to apply a holding time or the average length of time per peg count call. Representative holding times were difficult and costly to obtain. The usage equipment at the disposal of the traffic engineer to develop holding times was limited in capacity and subject to considerable error. Therefore, in the process of converting peg counts to usage, reliable holding times were at a premium.

It should also be mentioned that registers were read and the results summarized on a manual basis adding more inaccuracies and expense to the over-all problem. Clearly there was a need for an automatic traffic usage measuring device of proper capacity and utmost reliability.

The Traffic Usage Recorder—Description

In 1955 the Traffic Usage Recorder (TUR), developed by the Bell Laboratories and manufactured by the Western Electric Company, became available for service. Designed for a large capacity, a basic unit or frame is capable of testing up to 3,600 items of equipment almost simultaneously. The TUR is timed to scan all 3,600 test terminals every 100 seconds or 36 times an hour operating associated registers when busy equipment is encountered. In effect, the scorings indicate the number of calls of 100 seconds' duration, which is a basic unit of load measurement employed in traffic engineering. As an example, Table I[1] indicates the number of 100-second calls which can be handled by various size trunk groups to provide a grade of service equivalent to encountering one all-path-busy situation per 1,000 calls. For purposes of a brief general description, the TUR may be broken down into three components: the frame, register rack, and control panel.

The frame, pictured in Fig. 1, consists of timing units, scan and register crossbar switches, grouping fields, and control equipment. The timing units and control equipment contain the relay circuits, timers, and detector circuits that control the over-all functions of the TUR. Six scan switches provide the input terminations and busy signal paths from the items of equipment being measured to the detectors, as shown in Fig. 2. Inasmuch as the six scan switches are of the 100-point 6-wire type, 3,600 input leads may be accomplished. The register crossbar switches in turn provide the path from the detector for the reproduced busy signal. There are six register switches of the same capacity as the scan switches with the 3,600 terminals being cabled to the register grouping fields, where the register switch output terminals may be grouped to traffic registers as desired. Simultaneous operation of a scan switch and corresponding register switch crosspoints allows the inputs to a scan switch crosspoint (6 per crosspoint) to be tested, with any busy signals encountered being reproduced by detectors through the register switch crosspoint to the register grouping field. In addition to the register grouping field, there is a detector grouping field to associate the scan switch contacts with the proper detector.

The traffic registers associated with the TUR are located in a standard register rack bay shown in Fig. 3. Pin jacks at the base of the bay are cabled to punchings in the register grouping field located on the TUR frame. A pin jack may be connected to the desired register by means of a patch cord. Camera and hood assemblies are available to photograph the registers automatically, with two assemblies being required to photograph a rack of 300 registers.

The operation of the TUR is controlled at a relay rack or wall mounted panel consisting of a program timer, keys, and a cycle register; see Fig. 3. During automatic operation, the program timer controls scanning and minimum hour long scans may be programmed as required starting on the hour, half, or quarter hour. Also, the timer is equipped with a day wheel which will permit the elimination of any undesired day or days from the study period. The panel keys control manual scanning, cameras, tests, and miscellaneous operations. A total of 12 TUR frames may be served from one control panel.

Table I. Poisson Capacity Table

Trunks	Hundred Call Second Capacity	Per-Cent Usage
1	0.1	0.1
2	1.6	2.3
3	6.9	6.4
4	15.4	10.7
5	26.6	14.8
6	40.0	18.5
7	54.7	21.7
8	70.9	24.6
9	88.2	27.2
10	107.0	29 7

Paper 60-819, recommended by the AIEE Communication Switching Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting, Atlantic City, N. J., June 19–24, 1960. Manuscript submitted March 21, 1960; made available for printing April 22, 1960.

R. A. Bauerfeld is with the Southern New England Telephone Company, New Haven, Conn.

Fig. 1. Author and traffic usage recorder frame

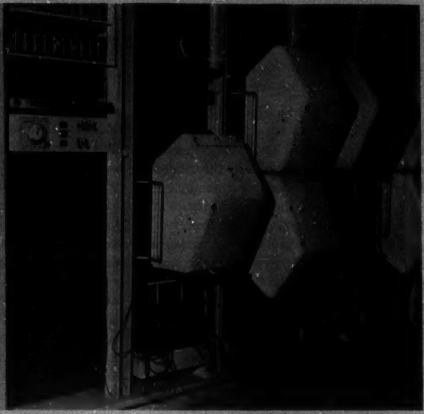


Fig. 3. Traffic usage recorder control panel and camera hood covered register racks

To insure the complete flexibility of the TUR, distributing frames have been provided at the Southern New England Telephone (SNET) Company TUR installations. The busy leads of equipment are cabled to one side of this frame, and the scan switch terminals to the other. This arrangement allows any item of equipment to be cross connected to any scan switch terminal.

The foregoing description of the TUR is by necessity brief as the intent is to provide an appreciation of the apparatus rather than a detailed understanding.

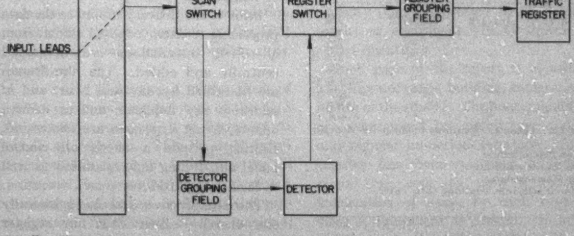


Fig. 2. Schematic showing interrelationship of traffic usage recorder components

The general review in the operation and capacity is considered essential to comprehend the possibilities and value of the TUR to the traffic engineer.[2]

Assignment of Equipment

The Southern New England Telephone Company, one of the first companies to adopt a comprehensive use of the TUR, had 55 frames in operation at the close of 1959, with 6 offices of over 10,000 connector terminals (telephone numbers) and all crossbar tandem offices are being equipped with the TUR; toll trunking is also being measured in the larger toll centers.

Equipment and trunks usually assigned to the TUR in local dial step-by-step offices include line finder groups, rotary and level hunting connectors, service code selectors, local and extended area trunking, and toll connecting trunks. Usage on the above equipment is essential to develop a complete distribution of traffic load through a local dial central office.

Trunk link frames and office link frames as well as major items of common control equipment are assigned to the TUR in crossbar tandem offices. Two TUR features, register grouping and 10-second scan, were not mentioned in the general description of the TUR, but are being used for measuring crossbar tandem equipment. The register grouping feature in a 20-trunk link and 20-office link frame office will allow the substantial saving of 300 No. 14 type registers. The 10-second scan arrangement is designed for equipment with exceedingly short holding times and will test for busy conditions every 10 seconds instead of the usual 100 seconds, lessening the possible error when measuring these items.

The entire Southern New England Telephone intra-company toll network is on a high-usage basis. Although this system is designed to realize trunking savings, a complete usage analysis is necessary to effect efficient loading and therefore all toll trunk groups, including intercompany trunks, are being measured in TUR equipped toll offices.

Scheduling

Studies designed to produce traffic engineering basic data are scheduled by the engineering group. The study period, or sample, should approximate the average busy season with reasonable accuracy and will depend upon the characteristics of the particular office involved. Therefore, study programs should be determined after a thorough analysis of data, and reviewed for accuracy before being applied. In general, SNET Company study periods may be discussed by the type of equipment and trunks being measured.

In step-by-step local dial offices, engineering data are obtained initially on the basis of 5 consecutive business days per month for an entire year. After the seasonal calling patterns have been determined, only months prior to or following an addition, and covering the busy season are studied. The study week is selected to exclude any influences, such as a holiday, in an attempt to obtain basic data during an "average" business week. Four hours of usage are obtained on each day during the study period. The 3:30 to 4:30 a.m. reading is designed to obtain an indication of the equipment which may be in trouble or busied out at the central office, and the 9:30 to 10:30 a.m. and 10.30 to 11:30 a.m readings to furnish a record of the morning busy hours. The fourth hour to be read varies from month to month. This hour is required to obtain the maximum usage of various types of equipment or trunks for which the busy hour does not occur in the morning period. With experience it may be possible to confine studies to a fewer number of hours, obviously reducing costs. All local dial central office TUR's are scheduled for the same study period permitting the comparison of data collected on an originating basis at various offices to be correlated in the development of terminating load distributions, thus avoiding the necessity of measuring an interoffice trunk at both ends. For purposes of illustration, a schedule of TUR readings in local dial step-by-step offices for the first half of 1960 is shown in Table II. Engineering of local dial offices is based on usage during the average busy hour during the busy season.

Unlike local SxS equipment, crossbar tandem equipment is engineered on an average of the "10 high day" busy hour usage base. Experience has shown that peak usage occurs on Mondays, therefore it is felt that a representative 10 high day base may be developed by analyzing Monday loads only. In addition, SNET Company traffic engineers undertake periodic studies to determine frame balance that will influence assignments of trunks.

Toll trunk group usage is desired for 5 consecutive business days per month. The particular week selected is determined in much the same manner as for local dial offices; however, the hours scanned are quite different. A total of 15 hours per day, 3 to 4 a.m. and 9 a.m. to 11 p.m., are scanned in order to develop group and network busy hours. This schedule is maintained 12 months per year. In addition, miscellaneous special studies are scheduled to resolve items such as length of day, a factor applied to busy hour usage to reflect total day usage, and toll traffic loads and patterns during periods of heavy calling such as Christmas or Mother's Day.

Benefits derived from the practice of formal advance scheduling include:

1. A periodic general review as to the basic data required to engineer equipment and facilities satisfactorily. This may lead to deletions or additions to the items of equipment being measured and/or revised study periods.
2. A base from which machine data processing work loads may be estimated.
3. A greater degree of TUR utilization by the local traffic administrator in regard to obtaining data for other than engineering purposes, such as administration of equipment or for special studies.

Processing of Data

Procedures have been developed by the SNET Company by which the automatically photographed TUR register readings are processed by IBM (International Business Machines Corporation) at the SNET data processing center at New Haven. Reports are produced indicating actual loading of the various types of equipment.

Prior to the delivery of film to the data processing center, exposed films from all TUR installations are developed centrally and edited. The film frames are identified by day and hour, and in addition any duplicate and/or camera operation test exposures are eliminated. Editing permits a check on control panel and timing unit operation as well as facilitating IBM key punch operation.

The data processing method is basically one whereby edited TUR film register readings are transferred by a key punch

Table II. Traffic Usage Recorder Schedule
Local Dial Offices

Month	Hours	Tabs Depressed	Month	Hours	Tabs Depressed
Jan. 1960.....	3:30– 4:30 a.m.	3:30 a.m.	.......Feb. 1960...	3:30– 4:30 a.m.	3:30 a.m
	9:30–10:30 a.m.	9:30 a.m.		9:30–10:30 a.m.	9:30 a.m.
	10:30–11:30 a.m.	10:30 a.m.		10:30–11:30 a.m.	10:30 a.m.
	6:30– 7:30 p.m.	6:30 p.m.		7:00– 8:00 p.m.	7:00 p.m.
March 1960..	3:30– 4:30 a.m.	3:30 a.m.	.......April, 1960..	3:30– 4:30 a.m.	3:30 a.m.
	9:30–10:30 a.m.	9:30 a.m.		9:30–10:30 a.m.	9:30 a.m.
	10:30–11:30 a.m.	10:30 a.m.		10:30–11:30 a.m.	10:30 a.m.
	12:00– 1:00 p.m.	12:00 p.m.		6:30– 7:30 p.m.	6:30 p.m.
May 1960....	3:30– 4:30 a.m.	3:30 a.m.	.......June, 1960..	3:30– 4:30 a.m.	3:30 a.m.
	9:30–10:30 a.m.	9:30 a.m.		9:30–10:30 a.m.	9:30 a.m.
	10:30–11:30 a.m.	10:30 a.m.		10:30–11:30 a.m.	10:30 a.m.
	7:00– 8:00 p.m.	7:00 p.m.		3:30– 4:30 p.m.	3:30 p.m.

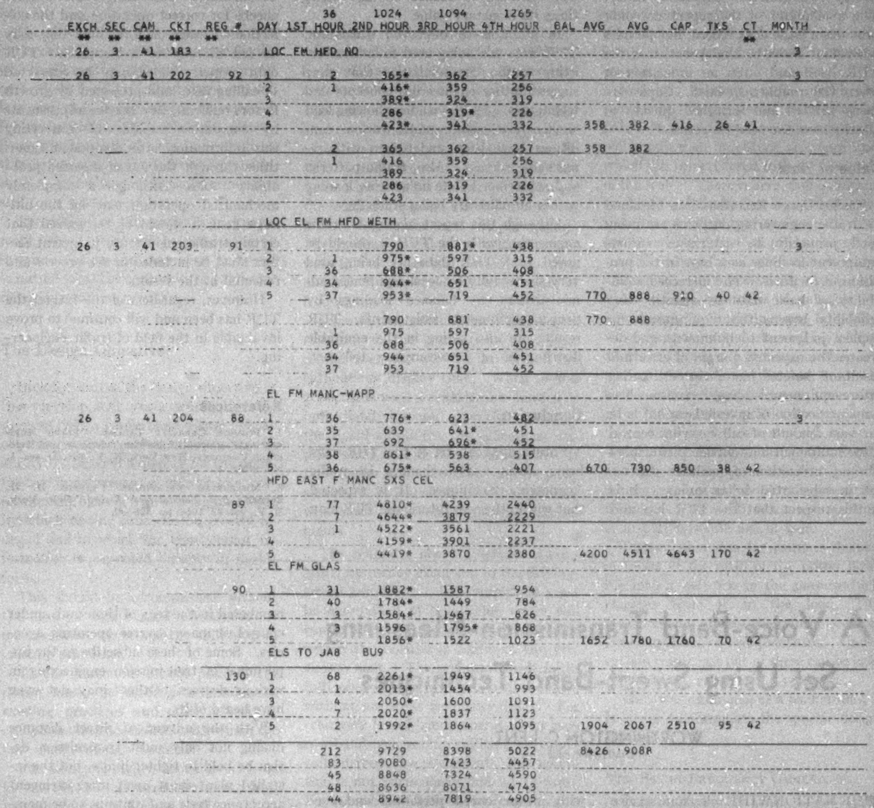
Detector Test: Following the reading scheduled for 11:30 a.m., on Friday, the Traffic Usage Recorder will be stopped, a detector test made and the results of the test photographed manually. This reading will replace the last scheduled Friday reading and will permit unloading of camera at the same hour each month.

Weeks to Be Read by District

District	Jan.	Feb.	Mar.	Apr.	May	June
Bridgeport (a)	1/4–1/8	2/1–2/5	3/7–3/11	4/4–4/8	5/9–5/13	NR
Hartford	"	"	"	"	"	NR
Manchester (b)(d)	"	"	"	"	"	NR
New Britain	"	"	"	"	"	NR
New Haven	"	"	"	"	"	NR
New London (c)	—	—	"	"	"	6/6–6/10
Norwalk	1/4–1/8	2/1–2/5	"	"	"	"
Stamford	"	"	"	"	"	NR
Waterbury	"	"	"	"	"	6/6–6/10

(a) Bpt. No. TUR completes 3/26/60, readings begin on Apr. schedule. Continue readings for Bpt. No. and Fairfield thru June.
(b) Manc. Main TUR completes 1–30–60, readings begin on Feb. schedule.
(c) N.L. Main TUR completes 2–20–60, readings begin on March schedule.
Nwch Main TUR completes 5–7–60, readings begin on May schedule.
(d) Torr. TUR completes 2–27–60, readings begin on March schedule, continue thru June.
NR—No Reading.

Issued: November 10, 1959
Office of the Dial Traffic Engineer



```
                                   36     1024     1094     1265
EXCH SEC CAM  CKT  REG # DAY 1ST HOUR 2ND HOUR 3RD HOUR 4TH HOUR  BAL AVG  AVG  CAP  TKS  CT MONTH
 **  **  **   **   **                                                                     **
 26   3  41   183        LOC FM HFD NO                                                          3

 26   3  41   202   92   1          2      365*     362      257                               3
                         2          1      416*     359      256
                         3                 389*     324      319
                         4                 286      319*     226
                         5                 423*     341      332    358    382  416   26  41

                                    2      365      362      257    358    382
                                    1      416      359      256
                                           389      324      319
                                           286      319      226
                                           423      341      332

                         LOC EL FM HFD WETH

 26   3  41   203   91   1                 790      881*     438                               3
                         2          1      975*     597      315
                         3         36      688*     506      408
                         4         34      944*     651      451
                         5         37      953*     719      452    770    888  910   40  42

                                           790      881      438    770    888
                                    1      975      597      315
                                   36      688      506      408
                                   34      944      651      451
                                   37      953      719      452

                         EL FM MANC-WAPP
 26   3  41   204   88   1         36      776*     623      482                               3
                         2         35      639      641*     451
                         3         37      692      696*     452
                         4         38      861*     538      515
                         5         36      675*     563      407    670    730  850   38  42
                         HFD EAST F MANC SXS CEL
                    89   1         77     4810*    4239     2440
                         2          7     4644*    3879     2229
                         3          1     4522*    3561     2221
                         4                4159*    3901     2237
                         5          6     4419*    3870     2380   4200   4511 4643  170  42
                         EL FM GLAS
                    90   1         31     1882*    1587      954
                         2         40     1784*    1449      784
                         3          3     1584*    1467      626
                         4          3     1596     1795*     868
                         5                1856*    1522     1023   1652   1780 1760   70  42
                         ELS TO JA8   BU9
                   130   1         68     2261*    1949     1146
                         2          1     2013*    1454      993
                         3          4     2050*    1600     1091
                         4          7     2020*    1837     1123
                         5          2     1992*    1864     1095   1904   2067 2510   95  42

                                  212     9729     8398     5022   8426   9088
                                   83     9080     7423     4457
                                   45     8848     7324     4590
                                   48     8636     8071     4743
                                   44     8942     7819     4905
```

Fig. 4. Reproduction of actual IBM traffic usage recorder results

machine operator into a working card deck containing identification codes gang punched from a master deck. The reading differences are computed and printed on a "report" illustrated as Fig. 4. In addition to differences, average and totals are also computed. From the completion of the key punch operation to the final report, all steps are accomplished by machine.

In order to develop a master deck of IBM cards, an essential tool of the data processing procedure, it is necessary to code all registers. This "coded" information is designed to: 1. Identify the report and the "working" card deck with a minimum of digits; 2. Aid in the programmed functioning of the machine; 3. Provide ready comparisons between results and engineering objectives. The coding for each TUR installation is maintained on a current basis by the traffic engineering group, and is identified on Fig. 4 by a double asterisk for reference. Two items are not coded, capacity (CAP) expressed in CCS, and trunks or equipment (TKS) shown as actual quantities.

Each page of the report is identified as to the exchange, building, camera, and month of the study. The hourly readings are summarized by circuit-type coding and register numerical sequence. Each register has been assigned a 3-digit circuit-type code to allow a machine totalization of usage by each type or class of equipment as desired. In addition to totals, the daily high hour, identified by machine with an asterisk, is averaged and printed under AVG. The column under BAL AVG is an average of the 2nd and 3rd hour or the 9:30 to 11:30 a.m. period, and is used to note the balance of certain items of equipment during the study period.

The general layout of the report facilitates a rapid evaluation of its validity. The results of the scan test, the 4th hour of the 5th day on Fig. 4, are easily compared to the quantity column substantiating the TUR circuitry from the detector to the register. A glance at the 1st hour, 3:30 to 4:30 a.m., will suggest permanents, troubles, or equipment out of service, and TUR trouble may be discovered by comparing a detector group usage register with a circuit-type total for the same entity. The availability of the report not only provides a basic data record, but permits a constant check on the operation of the TUR itself and often an indication of where the trouble may exist. Copies are easily printed and furnished quickly to all who have the need for results.

Value of Data

Traffic Usage Recorder data obtained for traffic engineering purposes are being used primarily to determine various equipment loadings as a base in the provision of facilities. The increased availability of basic data in quantity and reliability lessens the requirements for applied judgment to unknowns, and decreases the expected margin of error and resultant safety factor in estimating future equipment loadings. Inasmuch as a major portion of invested capital is in the vast amount of call carrying central office equipment and outside plant, more efficient utilization of facilities will result in substantial dollar savings. It is in this respect that the TUR has and will continue to realize savings many times its monetary value.

In addition to improved utilization, TUR data are being used to analyze existing traffic characteristics that may suggest future engineering concepts and techniques. Special studies showing load service relationships under various conditions have been undertaken with remarkable success. Also, traffic patterns and subscriber habits influencing loading within an office are being studied.

Although this report highlights traffic engineering with the TUR it should be noted that TUR data are being used very successfully by local equipment administrators to balance loadings by proper equipment assignments. TUR results are also aiding in the equitable distribution of intercompany toll revenues.

Conclusion

Future applications of the TUR may bring about many changes in traffic engineering techniques. It is expected that within the near future, the TUR may be modified permitting more direct checks for correct operation and the collection of usage on a medium that may be fed directly to a computor. TUR data properly coded may be converted to calling rates and projected by growth factors reflecting new service adjustments by computors capable of converting this information into equipment quantities through the use of digested probability tables. Although a completely mechanized operation may be the ultimate goal, it should be recognized that certain traffic engineering judgment factors must be included in the process and reflected in the result.

However, regardless of the future, the TUR has been and will continue to prove invaluable in the field of traffic engineering.

References

1. Poisson Capacity Tables. "Traffic Engineering Practices," American Telephone and Telegraph Company, New York, N. Y., div. D, sec. 1, p. 1 of fig. 1.

2. Measuring Telephone Traffic, D. H. Barnes. *Bell Laboratories Record*, New York, N. Y., vol. 32, 1954, pp. 326–30.

Vol.
Pt.
196

A Voice-Band Transmission-Measuring Set Using Swept-Band Techniques

WORTHINGTON C. LENT
NONMEMBER AIEE

THE NATIONWIDE telephone service which the country enjoys today is largely the result of careful systems engineering of which one objective is the establishment and application of uniform transmission standards. Transmission measurements constitute the essential means by which circuit design can be verified and the performance characteristics dictated by it maintained during operation.

The techniques for performing these measurements are well defined but tedious and time consuming. Achievement of useful results requires the effort of skilled personnel capable of exercising some judgment in application and interpretation. The plant has expanded at a rate exceeding that at which these skills can be supplied in sufficient quantity or at a cost which can be justified. There are only three alternatives: 1. to curtail testing, 2. to meet expanded testing loads with inexperienced personnel and correspondingly unreliable results, or 3. to revise techniques and equipment to reduce testing-time per circuit so that a limited number of properly qualified personnel can handle the increased load.

It is probably safe to say that, in general, the toll trunks have been carefully maintained. Their relative importance and the concentrations of equipment at toll terminals justify the assignment of skilled personnel to ensure proper operation. Furthermore, with the toll facilities under control of a single agency, the application of uniform practices and enforcement of maintenance procedures is possible.

In the exchange plant the situation is neither as simple nor as well defined. Where, in the case of the toll plant, hundreds of circuits under unified control might be involved, the exchange plant embraces trunks and subscriber loops numbered in the tens of thousands under control of many diverse operating agencies. Some of these subscribe to the importance of transmission engineering in various degrees. Others may not even have heard of it.

With the advent of direct distance dialing not only must transmission design be held to tighter limits, but the installed plant must meet more stringent acceptance tests and continue to be maintained with greater care than ever before. There will no longer be the operator to patch around deficient circuits. As a matter of fact, direct dialing, important and far-reaching as it is, could be a serious liability unless transmission design and maintenance practices adequate

---

Paper 60-816, recommended by the AIEE Wire Communication Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting, Atlantic City, N. J., June 19–24, 1960. Manuscript submitted March 24, 1960; made available for printing May 17, 1960.

Worthington C. Lent is with the Hallamore Electronics Company, Division of The Siegler Corporation, Anaheim, Calif.

The author wishes to express his appreciation for the co-operation and encouragement of Mr. Harold C. Diddel and Mr. Merle H. Jones of The Pacific Telephone and Telegraph Company at Los Angeles, Calif., and to acknowledge the very important contribution of Mrs. Eileen Kanczuzewski and Mrs. Marsha Rowan who typed the manuscript. Recognition is also due Mr. Arnold Schumacher and Mr. Robert Spaulding of Hallamore Electronics Company, who assisted during various phases of the development.

to support it are established and enforced.

The ultimate solution undoubtedly lies in the development of an automatic device which will continually probe every trunk on a whole-band basis for flatness and net loss and pass it for service or reject it for maintenance with a fault record. Some progress has already been made in this direction on a single-frequency basis. Such a device is certain to be expensive and some years away.

The subject of this paper is an instrument which may well be considered an interim solution and may even represent in nature what the economics of particular situations can ever justify.

The Design Objective

Broadly stated, the design objective of this development rests wholly on the premise that a whole-band visual display of transmission performance, with the ability to use such a display for the quantitative measurement of gain or loss, constitutes the basis of a measuring technique by which testing time can be greatly reduced and in which the transmission information is presented directly in usable form.

This should be accomplished without the need for complicated adjustments or interpretation exceeding the skill level of existing personnel.

To the maximum extent possible the device must be directly adaptable to existing practices and other measuring equipment and accessories with which the personnel is already familiar.

Portability is highly desirable, at least to the extent required for reasonable movement within a central office.

Specifically, the device must have the capabilities to:

1. Measure insertion gain or loss of any of the voice-band equipment or circuits now in service.
2. Provide a simultaneous whole-band display of relative input impedance, and the transmission-frequency characteristic at the output, of devices such as negative-impedance repeaters, equalizer networks, and hybrid networks, whose proper operation depends upon in-circuit adjustment.
3. Provide a whole-band display of the relative input-impedance characteristics of loaded lines to detect load misplacement or other impedance irregularities.
4. Display whole-band trans-hybrid transmission and measure the corresponding loss for the purpose of perfecting terminations and office balance.

There are undoubtedly many other applications of the device which will occur to the skilled worker.

The Design Problem

To reach these objectives several discrete circuit elements must be combined. Each will have performance requirements peculiar to it which must be specified.

Basically, any swept-band visual-display device must incorporate a generator capable of producing a constant-amplitude variable-frequency output over the desired band and a means for displaying the relationship which exists at a common instant between at least two parameters. The ordinary oscilloscope relates amplitude to time. Several devices present a display of amplitude as a function of frequency.

Quantitative information is derived by reference to the graduations of a scale placed over the viewing surface of the cathode-ray display tube. When such a graticule is used a considerable burden is immediately imposed upon the generator, the signal amplifiers, the deflections amplifiers and, indeed, the whole system. If this burden is not properly carried by the system, it is transferred to the operator who must make compensating adjustments.

The generator obeys a definite law relating frequency excursion to the driving-sweep amplitude. This amplitude must be maintained at that value which will cause precise registration of the trace within the horizontal limits of the prechosen scale. The generator must faithfully obey its own law, sweep after sweep, without apparent frequency jitter. For relatively broad-band generators, or narrow-band generators utilizing mechanically driven sweep elements, this condition is not particularly difficult to meet. For voice-band all-electronic generators it is.

Positioning circuits associated with cathode-ray devices are susceptible to drift unless elaborate regulation is incorporated.

The raw output of a swept-band generator appears on an amplitude-frequency display as a constant-amplitude phase-modulated wave. If this wave is passed through any device in which the transmission is not uniform with frequency, the amplitude of the envelope of the wave at the output of the device will reflect the variations in transmission. The envelope is not a trace but the locus of the peak excursions of the output wave. This means that the cathode-ray beam is spending most of the time in tracing out portions of the wave containing no useful information. The result is low trace intensity at the peaks because of the rapid writing-rate.

If some form of detection is used, the envelope information can be captured and the result is a well defined intense line, identical in nature to that of a manual graphic plot.

Now any arrangement by which the signal input to the display device can be reduced to zero during the retrace interval will produce an amplitude reference axis which can be registered with the base line of the scale. There is here, then, a means by which amplitude-frequency characteristics may be displayed and scaled. If it can be assumed that all of the necessary conditions have been satisfied and can be maintained, this system is capable of yielding large quantities of valuable information in a short time and in a form which permits visual observation of the effects of circuit adjustments as they are made. It must be emphasized, however, that it is bound to a prechosen scale with all of the attendant problems of registration in both amplitude and position.

Even in the best of the various forms in which it has appeared this arrangement remains only an amplitude-measuring technique. By itself, it cannot measure or display absolute loss or gain.

Transmission loss (or gain) is defined in terms of the ratio of the power flowing into a network to the power which emerges from it to the load. The measurement of loss, therefore, requires the comparison of two quantities. Under certain conditions one of these can be held constant and considered a reference.

With this discussion as a background it is easier to formulate the specific design problem.

The Swept-Frequency Generator

This vital element of the system must generate a sinusoidal wave of constant amplitude at all frequencies in the commonly defined voice-band, extending from 200–4,000 cps (cycles per second), without frequency discontinuities. The stability must be such that the band limits will be precisely reached on every sweep excursion and such that, within these limits, for every sweep the rate of change of frequency with time shall remain the same. This should be accomplished electronically without resort to mechanical devices.

The output level should be adjustable from zero to the maximum employed in the plant, usually about + 10 dbm (decibels referred to 1 milliwatt).

There are two broad classes of measurements, each of which requires a different internal generator impedance. All insertion-loss measurements are made between specified impedance for both the genera-

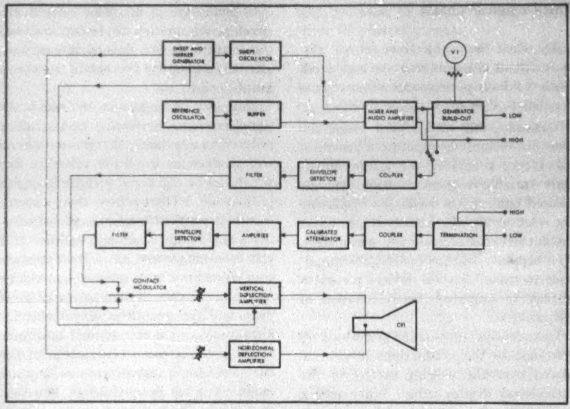


Fig. 1. Functional block schematic

tor and the load. Common practice has emphasized the use of both 600 and 900 ohms. Others may be of interest. Since the impedances may range, the generator itself should have an output branch of very low internal impedance which can be built out to meet specific measurement conditions.

Some apparatus, such as negative-impedance repeaters, depend upon in-circuit impedance conditions for proper operation and can most effectively be adjusted under those conditions. Any measurement device must not disturb them. This can be assured by providing the generator with a high-impedance output branch which will permit bridging at the line terminals.

Since the circuits and the access points of most of the voice-band equipment in the plant are balanced to ground, the generator output branches must also be balanced.

THE DISPLAY

In general, the display must portray, in visual form, the information which the measurement system produces. It must be as large as reasonable size and circuit complexity will permit so that small changes in trace position may easily be resolved. The trace persistence must be such that all of the information presented during at least one complete sweep interval will be visible as a whole. Intensity and trace resolution must permit permanent photographic records to be made with standard equipment. The associated positioning circuits and controls must allow movement of the trace over the entire viewing surface. Deflection in the vertical direction should be related to signal amplitude. Horizontal deflection should be related to the frequency excursion of the swept-frequency generator with the low frequency limit of the band appearing at the left.

THE MEASUREMENT SYSTEM

This, of course, is the heart of the instrument. If it is to measure the absolute loss (or gain) of a circuit or device, provision must be made to take the input power as a reference, compare the power emerging from the device and present the comparison visually in such a way that it can be quantitatively evaluated. The condition for measurement should be completely independent of the absolute level at which the measurement is made. Above all, the quantitative evaluation should not depend upon a graticule, etched scale, or any other graphic scaling arrangement with their associated difficulties of registration and control and other inherent limitations. One condition, and one condition only, should be necessary and sufficient for measurement and the readout should be simple and without ambiguity.

THE SWEEP-DRIVE SYSTEM

The sweep-drive system embraces those portions of the instrument in which all of the timing functions are performed. The basic waveform used to drive the swept-frequency generator over the required range must be generated and shaped in a manner which will result in the proper distribution in frequency along the horizontal axis of the display.

Since no pregraduated scale is to be used with the display, it is necessary to generate a series of marks which, when superimposed on the trace, will permit location of discrete frequencies.

The choice of sweep-rate is related to the persistance of phosphors available in cathode-ray tubes and to the time required for retrace transients to decay to the level below that at which confusion in the display might occur.

ENVELOPE DETECTION

If the display is to contain only those portions of the output wave which bear useful information, some form of envelope detection is necessary. The time-constant of the output filter is important and must be carefully chosen. The upper limit for the cut-off frequency is determined by the lowest frequency in the output wave at which it is desired to display only the envelope. Since, in this case, the lower band-limit has been fixed at approximately 200 cycles, the cut-off frequency must be well below that to avoid clutter from unfiltered components.

The lower limit is fixed by the degree to which the device must accurately portray circuit conditions. One of the most valuable applications of this instrument is the visual display of the standing-wave effects produced on lines by impedance irregularites due to mistermination or improper loading. The filter cut-off frequency must be sufficiently high to permit these effects to be displayed without restriction of their amplitude.

The Design Solution—Functional Description

Having explored and enumerated the several design problems it is now appropriate to examine the means by which the solutions to them have been achieved.

Reference to Fig. 1 will be helpful for this purpose. The various portions of the system will be discussed in the logical order established by the function which each performs.

THE SWEEP AND MARKER GENERATOR

The basic timing waveform is generated by an astable multivibrator operating at a rate of 5 cps. The square-wave output is used for two distinct purposes; to provide pulse marks on the trace for the location of discrete frequency points on the horizontal axis of the display, and to control the timing of the discharge of condensers used in the production of a sawtooth waveform. This timing is such that 1 cycle of the sawtooth is generated for every 4 cycles of the multivibrator.

The sawtooth is used to deflect the

cathode-ray beam in a horizontal direction and simultaneously to effect the frequency excursion of the swept frequency generator. It is produced by driving a step-integrator with pulses from the multivibrator. The grid of a relay tube, in whose plate circuit is inserted the coil of a mechanical relay, is connected to the integrating condenser and so biased that plate conduction occurs on the fourth step of the integrator output wave. The resulting plate current is sufficient to close the relay contacts. There are two sets of contacts. One set is connected across the integrating condenser; the other is connected across a condenser which is charged through a suitable resistor from the main power supply. The values of capacitance and resistance are such as to produce a linear rise of voltage (during the time before the relay operates) which forms the ramp of the sawtooth waveform. When the relay operates both condensers are discharged abruptly and completely and the process repeats.

The synchronous operation of two separate systems is an important requirement for certain types of transmission measurements. Such operation requires rigid control of both the frequency and phase of the individual timing generators. Frequency control is achieved by using the 60-cycle power source as the reference standard. A 12 to 1 countdown circuit generates synchronizing pulses which are applied to the multivibrator whose rate is accurately locked at 5 cps. Phase control is achieved by providing a push-button switch by which the sawtooth generator can be stopped and restarted at any instant.

THE SWEPT-FREQUENCY GENERATOR

There seems to be no direct way, except through the use of mechanically driven elements, to generate phase-modulated waves in the voice-frequency range. Beat-frequency methods are well known and have been used extensively in fixed frequency, and to some extent in swept-band, equipment.

The usual approach is to use a reactance tube as the modulating element in a variable oscillator referenced against a second oscillator operating at a fixed frequency such that the difference is zero when the reactance tube is not excited. The fact that both the variable oscillator and the modulating elements must effect changes in frequency implies instability. This becomes evident both as jitter and in the need for frequent adjustment to compensate for drift.

Since the low-frequency limit of the voice-band is in the region of 200 cps. jitter and drift must be minimized to avoid confusion in the display and error in the presentation in frequency.

The swept-band generaor used in this development is believed to be an entirely new means for generating phase-modulated waves in the audio-frequency range with stability in excess of that now possible by any other known all-electronic methods.

Semiconductor development has recently produced voltage-sensitive elements whose terminal capacitance is a function of the applied reverse-bias voltage. These have been used in many ingenious ways as variable-capacitance devices. When used to vary the frequency of an ordinary oscillator it is necessary, because of their limited capacitance range, to choose a relatively high median frequency or to tolerate a highly instable oscillator at a lower frequncy, if appreciable frequency excursion is required. Neither alternative yields the desired degree of stability.

It is well known, however, that quartz crystals have proved to be outstanding as highly stable elements for the control of frequency. A crystal has two limiting frequencies. The series-resonant frequency is determined solely by the internal crystal constants and cannot easily be altered except by physical modification of the crystal. The antiresonant frequency is determined by the combined effect of the internal constants and the terminal to terminal capacitance which is composed of the capacitance of the crystal mount and whatever capacitance is introduced by the holder and the external circuit. The limiting antiresonant frequency will be that at which all external shunt capacitance is zero. Certain cuts are capable of yielding a range between the limiting frequencies of about 0.04% of the series-resonant frequency.

If a suitable crystal is used in a circuit employing the antiresonant mode, it is possible to vary the frequency by varying the shunt capacitance. The voltage-variable diode is an ideal element for this purpose since it, like the crystal, is temperature-stable.

The crystal frequency cannot be varied over the entire range established by the limiting frequencies. As the series-resonant limit is approached the impedance decreases rather rapidly to the point where the voltage necessary to support oscillation can no longer be maintained.

The generator incorporated in this instrument uses two crystal oscillators operating at approximately 19 mc (megacycles). One of these can be varied continuously in frequency over a range of approximately 5 kc by use of voltage-variable diodes properly biased. The sawtooth waveform from the sweep-drive generator is shaped and superimposed on the bias to provide the cyclical frequency excursion.

The oscillator outputs are fed through load-isolating buffer stages to a modulator in the output circuit of which the phase-modulated wave appears. The oscillators and the sweep drive are adjusted so that the difference frequency ranges from 0–4,000 cps.

The output of the mixer is amplified and fed, in split phase, to a balanced output stage. In order to meet the impedance requirements of the various types of measurement, two branches are provided. The low-impedance leg has an internal impedance of approximately 30 ohms. Jacks are provided for access to internal buildout at both 600- and 900-ohm impedance levels for use in the more common insertion-loss measurements.

The high inpedance leg has an internal impedance of approximately 25,000 ohms which permits bridging the generator across lines without disturbing their impedance conditions.

All access to the generator is on a balanced basis so that no auxiliary transformers are required for any of the normal applications in the telephone plant. Level is adjustable from zero to +10 dbm.

THE MEASUREMENT SYSTEM

The prime requisite of the measurement system is to evaluate a power-ratio independent of the absolute power level. This requires a reference point which is invariant in both amplitude and impedance. Such a reference point is the low-impedance generator output branch. Inasmuch as the impedance is about 30 ohms, any normal buildout will provide sufficient isolation from the load to avoid disturbance. Essentially, then, this is a reference point of constant power. Proper equalization can establish this constancy for all frequencies within the swept band. A voltage-sensitive device is sufficient for indicating the reference condition.

The insertion loss of any device is defined as the ratio of the power which would appear in a specified load, when connected to a generator having an internal impedance equal to the load, to the power which would appear in the same load when the device to be measured is inserted between the load and the generator. The voltage across the load without the device inserted can be considered a reference for power at the load and as

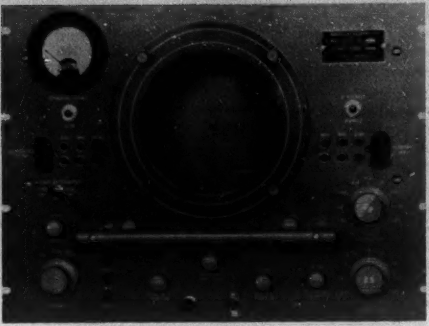


Fig. 2. Front panel view

long as the voltage is held constant the power will be also.

When the device is inserted, the voltage across the load will change. The power will vary as the square of the voltage so that some special type of indicator would be necessary to measure the changes in power.

Suppose, however, that an amplifier is also inserted and that the gain can be adjusted exactly to the value which nullifies the loss of the device under test. The voltage across the load, and therefore the power, will be the same as before the insertion and the indicator will read the same. The need for special characteristics is removed. In fact, not even a calibrated scale is required.

Now, instead of making the gain adjustable, a fixed gain is inserted with an attenuator whose loss can be adjusted. Attenuators can be highly accurate and stable in calibration. With this arrangement, it is possible to make loss measurements completely independent of the absolute voltage level. To do this, the amplifier, attenuator, and indicator are connected to the load before the device under test is inserted. The attenuator is adjusted to a reference setting which will cause a convenient indication. The device is then inserted between the generator and the load and the attenuator readjusted until the same indication is obtained. The difference in the attenuator settings is the loss of the device. Most of the transmission measuring sets in common use today operate in this manner.

The same technique is used in this new instrument. The system employs two independent channels, one of which is associated with the generator for the purpose of establishing the input-power reference conditions. The other branch, which includes the calibrated attenuator and amplifier is used to detect and evaluate changes in output power. Both branches terminate in identical envelope detectors, associated with which are low-pass filters which pass only the d-c and very-low-frequency components in the detector output.

A contact modulator, or "chopper," is used to switch the detector outputs alternately to the input of the vertical deflection amplifier associated with the display. Two distinct traces are thereby produced, one representing the input power, which becomes the reference trace, and the other portraying the envelope of the voltage appearing across the load. The d-c component is preserved through the deflection amplifier so that the traces remain in the absolute positions established by the d-c conditions only and are completely independent of each other. The reference trace, therefore, remains in the set-up position on the display, leaving the other to move freely in relation to it.

The reference channel is identified as the *A* channel, while the channel used for comparison is called the *B* channel. Balanced-to-unbalanced couplers, having very high input impedance, are used to connect the two channels in the various ways dictated by the requirements of the measurement conditions.

The *A*-channel coupler is always associated with the generator. Switching is provided to transfer the coupler input to either the low- or high-impedance generator output branch.

The *B*-channel coupler input appears at jacks, some of which are arranged to pick up internal terminations of either 600 or 900 ohms.

Calibration of the *B* channel is accomplished in all cases by patching the *B* coupler, with a chosen termination, to the corresponding generator output. With the attenuator at reference zero, adjustment of *B* channel gain establishes coincidence of the two traces. This gain adjustment is not distributed during measurement except for recalibration, should it be necessary. All measurements are made, thereafter, by simple attenua-

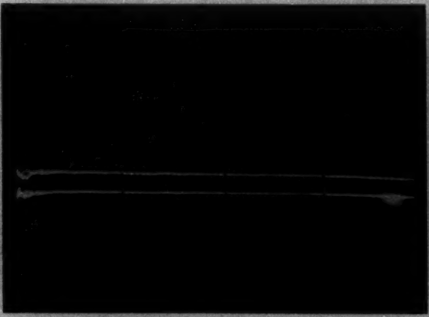


Fig. 3. Typical display under calibration conditions with B-channel trace depressed 1 db from coincidence

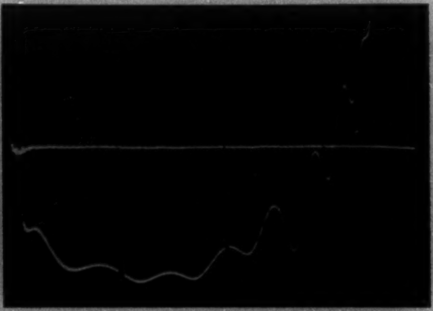


Fig. 4. Typical display of trans-hybrid transmission. Trace coincidences occur at 150 and 3,200 cycles, representing loss of 15 db

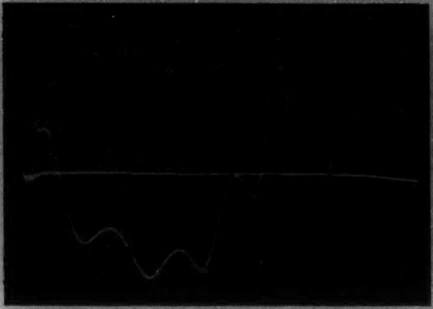


Fig. 5. Display of trans-hybrid transmission with cross-over points at 350 and 2,300 cycles, corresponding to a loss of 25 db. Portions of the B traces remaining below the A trace represent higher loss

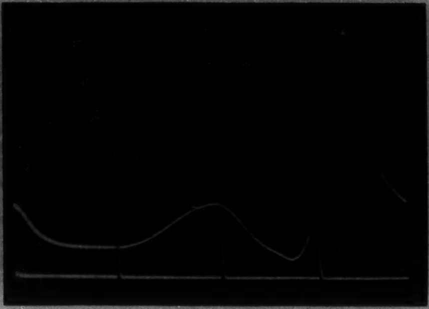


Fig. 6. Display of relative impedance at the input of an artificial 22-gage H-88 loaded line with third load removed

tor manipulation to restore trace coincidence.

For insertion measurements, the *B* coupler patch is removed and transferred to the output of the device under test. The built-out generator is, of course, patched to the input of the device. Any loss or gain in the device is evidenced by the departure of the *B* trace from coincidence with the *A* trace. At those points in frequency, where coincidence is retained the gain or loss is zero. Coincidence may be achieved at any frequency within the displayed band by manipulating the attenuator. The gain or loss of the device is read directly from the attenuator dials to an accuracy of 1 db and may be estimated visually well within that limit. Sufficient gain and attenuator range are provided so that losses as high as 55 db and gains as high as 35 db may be measured. This range is sufficient to embrace that required for measurement of all normal trans-hybrid loss and the gain of all commonly used repeaters.

For bridging measurements, such as those at the terminals of negative-impedance repeaters, the high-impedance generator branch is used. *B* channel calibration is established by patching the *B* coupler, without termination, to the generator. All other steps in the measurement procedure remain the same as for insertion measurements. The reference trace, will, however, now be sensitive to the impedance variations at the repeater and will, therefore, yield important information as to the ability of the strapping to provide nonreflective terminations for the lines.

In connection with measurements of the gain of negative-impedance devices, in circuit, there are certain limitations which should be emphasized. The test sets now in existence sample the line currents flowing into and out of the device and convert them to voltages for measurement. The gain thus measured is absolute only when the impedance on both sides are identical. The apparent gain, as measured, will deviate from the actual insertion gain to the degree that the impedances depart from equality.

This new instrument, while indicating directly the voltages across the lines instead of the currents flowing in them, is, nevertheless, subject to the same limitations since a true power ratio cannot be evaluated from voltage or current measurements alone unless it is known that the impedances remain alike. This is the reason for the discrepancies which have been noticed between insertion gains and those measured directly at the repeater. Usually, the strapping is successful in establishing impedance conditions which minimize the discrepancy.

The location, in frequency, along the horizontal axis of the display is established by pulse-type marks appearing at 1,000-cycle intervals. The start of the trace is, of course, zero and the generator frequency excursion is adjusted so that the termination of the trace occurs at 4,000 cps. Three distinct marks will appear in the interval, at 1,000, 2,000, and 3,000 cps. Since the sweep excursion is essentially linear, visual interpolation between the marks is relatively easy and the points of trace coincidence may, therefore, be located in frequency.

THE DISPLAY

The display system makes use of a 7-inch cathode-ray tube with a long-persistence *P7*, phosphor. An amber filter is placed over the tube face to accentuate the persistence characteristics and to reduce the intensity of the blue-phosphor response.

The tube is electrostatically deflected and focused. The deflection amplifiers are direct-coupled throughout, in the horizontal to preserve sweep linearity and in the vertical, to retain the d-c and low-frequency information in the signal.

Intensity, focus, and positioning controls are provided. Operation potentials are obtained from a high-frequency supply.

With the normal intensity which yields maximum spot resolution, good, printable negatives may be made using Panchromatic-X film, with an aperture of 12.5 and a 6-second exposure. Full-frame views of the tube face will require some sort of lens extension or copying attachment for ordinary cameras. A camera mount will be available as an accessory.

ADDITIONAL CIRCUIT FEATURES

Several convenient circuit features have been incorporated.

The most important of these is to provide the means for determining that the sweep excursion represents exactly the range from 0–4,000 cps, since the marks are uniform in time and must, therefore delineate 1,000-cycle intervals along the horizontal axis of the display. Two networks are used for this purpose and are available on a switched basis to establish the proper alignment. One of these produces a sharp notch at 1,000 cycles and a front panel control on the reference oscillator is used to align this notch with the first pulse-mark. It has been determined that this process insures that the left-hand limit of the trace corresponds to zero frequency. The other network produces a sharp notch at 4,000 cycles and another front-panel control is used to adjust the sweep-oscillator drive until the notch coincides with the right hand limit of the trace. The two remaining pulse marks then accurately locate the 2,000- and 3,000-cycle positions in frequency.

Inasmuch as two independent traces appear on the display, it is desirable to be able to identify one from the other. For this purpose, a push-button switch is provided to disable the *B* channel. A monitor jack is provided so that the operator may listen to the *B*-channel output.

It is contemplated that this instrument will ultimately be used for many of the now-standard purposes in the telephone plant. Presently, it is possible for the installer to dial from an outlying point on a circuit and cause a 1,000-cycle test tone at 0 dbm to be impressed on the office end of that circuit. Provision is made in this instrument to meter the output when the 600-ohm build-out is used so that a swept-band output at 0 dbm may be supplied on the same basis. A companion instrument can then be used by the installer to test the circuit over the entire band. Single frequency tests can often lead to false conclusions as to circuit performance.

Conclusion

Several photographs have been provided to show the physical appearance of the instrument and typical displays under actual measurement conditions.

Fig. 2 shows the instrument as it will appear mounted in a standard equipment bay. Fig. 3 shows the display as it will appear under calibrating conditions with the *B*-channel trace (lower) depressed one db from coincidence. Fig. 4 shows a typical display of trans-hybrid transmission. The reference trace is the straight line. The attenuator has been adjusted so that the trace coincidences occur at 150 and 3,200 cycles and represent a loss at those frequencies of 15 db. Fig. 5 is a display of the same conditions as in the preceding figure but with the attenuator loss decreased 10 db. Here the crossover points occur at 350 and 2,300 cycles and correspond to a loss of 25 db. All portions of the *B* trace remaining below the *A* trace represent even higher loss.

Fig. 6 is a display showing the variation of relative impedance at the input end of an artificial H-88-loaded 22-gage line, with the third load removed.

High-Voltage Static Eliminators for Teletypewriters

L. A. JONAS
NONMEMBER AIEE

F. E. HANSCOM
NONMEMBER AIEE

HIGHLY unfavorable operating conditions are often created by the depositing of static charges on teletypewriter paper tape and rolls. This phenomenon of static electrification will cause delays and require the attention of operating personnel. Messages can even be lost as the copy may jam in moving mechanisms and be mutilated. Although radium devices have been employed for a number of years by the Bell System for static control,[1] the increasing number of state laws regulating the use of radioactive substances made it desirable to find some suitable substitution. The paper discusses the operating principles and design features of a static eliminator utilizing a high-voltage discharge.

Static Factors

In teletypewriter equipment, paper undergoes a number of pressure contacts with other materials. Such contacts include sensing and punching pins, type pallets, platens, sliding surfaces, and the like. The paper can, by some such process, acquire a charge on its surface. The quantity and decay time of the charge are governed by a number of variable factors including paper composition, types of opposing surfaces, relative humidity of the ambient air, moisture content of the paper, and the amount of pressure applied. The influencing factors and the electrostatic phenomenon itself are complex, and for the sake of brevity this background material will not be covered here. It should be realized however that such information is essential for the proper design and evaluation of any static control arrangement. For those desiring more information the list of references is included.

Theory

If the air surrounding charged material is ionized, the material can neutralize itself by selecting those charges it needs to balance its surface electrically. This requires that a sufficient number of charges of the opposite polarity be available. One means to accomplish ionization of ambient air is to apply a high voltage between point electrodes and nearby grounded members. Any free electrons at the negative end will be repulsed toward the positive side. Experiments have shown that there are always some free electrons in the air. These electrons may have been produced by the action of light or cosmic rays, or as a result of the collision of molecules. As the electrons are sped through the air they collide with neutral atoms and displace electrons from the outer ring. These electrons will in turn displace others. By the utilization of alternating voltages both positive and negative ions can be supplied, since each in turn would be repulsed by the discharge points.

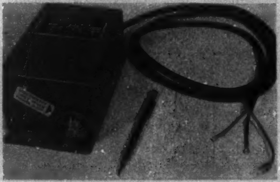


Fig. 1. Power unit

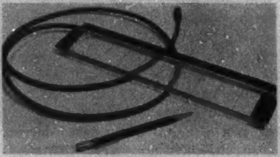


Fig. 2. Inductor bar

Paper 60-789, recommended by the AIEE Telegraph Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting, Atlantic City, N. J., June 19–24, 1960. Manuscript submitted March 17, 1960; made available for printing May 4, 1960.

L. A. Jonas is with the Bell Telephone Laboratories, Inc., New York, N. Y., and F. E. Hanscom is with The Portland Company, Portland, Maine.

Vol.
Pt. 1
196

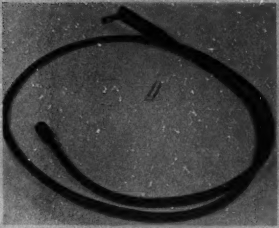


Fig. 3. Inductor point

Utilizing sharp points greatly reduces the amount of voltage necessary to achieve the beginning of ionization, because the electric field is most intense at narrow points. Although an electric field is always perpendicular to a conductor, it is not necessarily uniform but is a function of the geometrical shape of the conductor and will tend to concentrate in those places where the radius of curvature is least.

Application

The static eliminators have been designed as three separate units:

1. A power unit to supply the required voltage at a safe current level; see Fig. 1.
2. An inductor bar for mounting on the model *28* and model *15* page printers; see Fig. 2.
3. An inductor point suitable for mounting on the following tape equipment; see Fig. 3:

Model No.	Equipment
28	typing reperforator
28	reperforator transmitter
14	typing repreforator
14	reperforator transmitter
14	typing unit
28	automatic sending-receiving set
19	automatic sending-receiving set

Power Unit

The power unit consists of a 5,000-volt transformer with a capacitor. These are sealed with a potting compound in a 4-inch by 6-inch by 3-inch metal case. The primary is designed for 117 volts, 50 or 60 cycles. A capacitor has been installed in series with one side of the secondary winding to limit the output current to a maximum of 1.4 milliamperes. This side terminates at a duplex outlet on one end of the case. The other side of the secondary is connected to both a ground stud on the case and to the grounding wire of a 3-wire 6-foot power input cord.

An identical power supply is used for both the inductor bar and the inductor point. The duplex outlet provides for simultaneous use of two static eliminators as in the case of the model *19* or *28* automatic sending-receiving sets, where it may be required to control static on both tape and page copy. Also, there are installations where a large number of table console models of teletypewriters are installed adjacent to and on top of each other. In such cases the savings involved in reducing the purchases of power units could be considerable. In the event that one outlet is used, a plug is supplied as a seal for the unused outlet.

Two flanges with 1/4-inch holes can be used for mounting. The unit may be placed on a shelf or fastened to a rack or wall. The only requirement in locating the power unit is that it be within reach of the inductor cable, out of the customer's or maintenance man's way, and in a dry area.

Inductor Bar

Preliminary studies were conducted to determine the quantity of the static charge on teletypewriter page copy and the best location for a static eliminator. By the use of an electrostatic voltmeter several charge-creating areas in the model *28* teletypewriter were investigated. It was found that at least 70% of the total instrument reading was obtained by the action of the paper being compressed by the rubber platen, and that almost all the remaining reading was obtained by the action of the steel-type pallets striking the paper.

The effects of unrolling the paper, rubbing against the glass and metal edges and sliding over the dome, appears to be negligible. Therefore it was felt that treating the paper immediately after it left the typing area would be sufficient.

Another requirement was that there be some air space around the section of the paper being treated. If the paper is in contact with any solid surface, both the paper and the contacting body will be neutralized as a whole, but once the paper is parted it will be unbalanced electrically as a separate body.

Due to physical interference it was necessary to locate the inductor bar in front of the typing unit and under the glass viewing cover. Accordingly, a transparent bar with sharp points was constructed. There are nine recessed discharge points spaced approximately one inch apart to give full coverage to the paper. The material used is Plexiglass no. *55* which has stability and other required qualifications. A wax finish is given in production to afford protection against marring the surface.

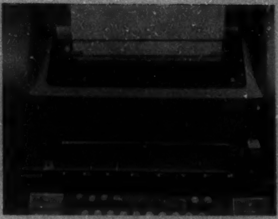


Fig. 4. Inductor bar on model 28 teletypewriter

Sliding metal brackets are provided at each end of the bar. They are adjusted with the aid of a screwdriver to fit either the model *28*, Fig. 4, or the model *15* teletypewriter page printers. One side of the bracket is painted black to reduce glare to the teletypewriter operators.

A bare wire is stretched along the length of the bar a short distance in back of the points. The wire is bonded to the metal brackets which in turn are grounded to the teletypewriter cover when they are installed. Without such a close proximity ground the intensity of the electric field would be greatly attenuated and a much higher voltage would have to be employed, probably in the neighborhood of 15,000 volts rather than the present 5,000 volts.[2]

A return path is provided from the proximity ground wire through the frame of the machine to a common ground with the power unit. In most cases the input for the power unit and the teletypewriter will be the same, and this can be used to provide the common ground. If such is not the case the ground stud on the power unit should be bonded to the teletypewriter frame.

Every point in the bar has its own capacitor. This coupling is utilized to restrict further the current to such an infinitesimal amount that not even a mild shock can be felt by touching the points. Measurements show a maximum current to the inductor bar points of less than 0.05 milliampere.

A rubber-covered high-voltage cable is permanently fastened to one end of the bar and a screw-type connector is fastened at the other end of the cable to permit connection to the power unit. There is some power loss in the cable due to inductive effects but it was felt that the use of stand-off insulators was an unnecessary complication. To keep this inductive loss to a minimum, however, the cable is restricted to a maximum length of 5 feet.

Tests were carried out in which the

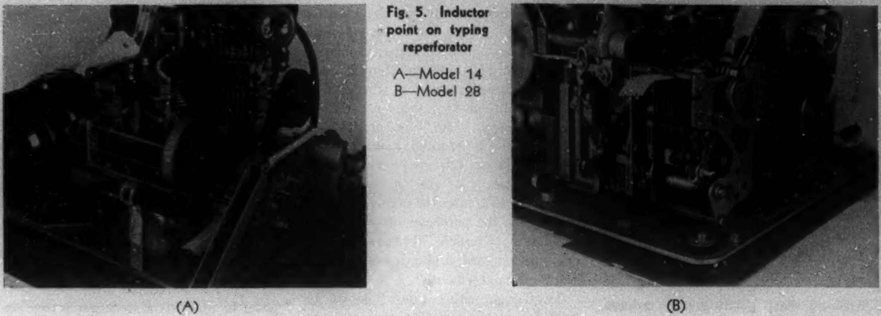

(A) (B)

Fig. 5. Inductor point on typing reperforator

A—Model 14
B—Model 28

distance from the static bar points to a charged web was varied, and the resulting voltages were measured. They indicated that the optimum distance would appear to be slightly more than one inch. That direct adjacence to the paper is not necessary is an advantage in so far as freedom for locating the neutralizer is concerned.

On the model *15* teletypewriter the distance from the discharge points to the paper is approximately one inch. On the model *28* the distance is somewhat less. This compromise is necessary because one line of the type message is obscured due to the nontransparent edge of the inductor bar. By positioning the bar this line of type is approximately four to five lines above the immediate typing area. It is felt that this would be a very minor objection.

In field and laboratory trials involving the model *28* page printers, static potentials of over 10,000 volts were repeatedly reduced virtually to a zero level by the use of the static bar. How low the actual potential must be kept on teletypewriter paper to avoid difficulties is not yet fully determined. Some tests appear to point to a potential gradient below 1,000 volts. There is also the possibility that further charges might be accumulated in processing after the paper has been neutralized. Therefore it would be advantageous to have the charge remaining after neutralization as low as possible.

Inductor Point

The inductor point consists of a single discharge point mounted at right angles to a coupling capacitor body. As in the case of the bar, the point is recessed. Recessing not only affords protection against scratching of the skin in handling but also lessens the possibility of arcing due to the point being too close to a ground. Arcing not only decreases the efficiency of the static eliminators but can also cause radio interference. The inductor point is made of a highly shock-resistant plastic. This is necessary because of its fairly exposed mounting. A simple 2-piece clamping device is supplied to arrange fastening on any of the tape machines previously listed. It is equipped with a 5-foot inductor cable, as in the case of the inductor bar, to connect to the power unit. No close proximity ground is provided on the inductor point itself since the teletype equipment is able to perform that function. A common ground with the power unit must be assured as with the inductor bar.

Two of the proposed mounting arrangements for the inductor point are illustrated in Fig. 5.

Ozone and Corona

Ionization of air is accompanied by the discharge of ozone gas. Ozone has a fresh, penetrating odor resembling that of very dilute chlorine. The action is reversible and after a little ozone has been processed it is decomposed at the same rate that it is generated. Ozone can be injurious in high concentrations. The maximum allowable concentration in air is given as 0.1 to 1.0 part per millionth of air.[3] The inductor bar was tested in the laboratories with an ozone meter capable of registering amounts far below those of even irritating concentrations. The quantity of ozone liberated by the designed units were so minute as to be undetected on the instrument.

Ionized air at the surface of conductors appears as small tufts and streams visible in the dark. Such ionized air is called corona because of its resemblance to solar corona. With alternating current, corona consists of a more or less continuous glow, reddish blue in color. This effect is visible only in full darkness.

Summary

The high-voltage static eliminators, as designed for teletypewriters, have certain distinct advantages. Unlike radioactive elements, they present no safety hazards. The inductor bar, inductor point, and power unit have been approved by the Underwriters Laboratories, Inc., and the Canadian Standards Association Laboratory; field tests indicate a high ionizing potential which is sufficient, even in severe static conditions, to reduce the static charge virtually to zero. The units are adaptable for a large number of teletypewriter models and surroundings with a minimum of effort. They can easily be modified for those installations where standard mounting arrangements are not yet available. The equipment has a wide positioning range so no fine adjustments are required.

References

1. STATIC ELIMINATION ON BELL SYSTEM TELETYPEWRITERS, **H. Bayley, D. B. Perry, B. S. Swezey.** *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 75, Mar. 1956, pp. 92–94.

2. TESTS OF ELECTROSTATIC CONTROLS FOR HAZARDOUS INDUSTRIAL APPLICATIONS, **Robin Beach.** *Ibid.*, pt. II (*Applications and Industry*), vol. 71, Jan. 1952, pp. 1–13.

3. HANDBOOK OF DANGEROUS MATERIALS (book), **N. I. Sax.** Reinhold Publishing Company, New York, N. Y., 1951.

4. A METHOD FOR STUDYING DISTRIBUTION AND SIGN OF STATIC CHARGES ON SOLID MATERIALS, **H. H. Hull.** *Journal of Applied Physics*, New York, N. Y., vol. 20, Dec. 1949, pp. 1157–59.

5. THE ROLE OF ASYMMETRIC RUBBING IN THE GENERATION OF STATIC ELECTRICITY, **P. S. Henry.** "Static Electrification," *British Journal of Applied Physics*, London, England, supplement no. 2, Mar. 1953.

6. METHODS OF INCREASING THE ELECTRICAL CONDUCTIVITY OF SURFACES, **J. S. Forrest.** *Ibid.*

7. ELECTRICAL CHARGE PHENOMENA IN PAPERS, **W. Brecht, F. Schmid, R. Vieweg.** *Papier-Fabricant*, Darmstadt, Germany, vol. 35, technical part, Mar. 1937, pp. 133–40, 142–59.

8. THE CONTACT ELECTRICITY OF SOLID DIELECTRICS, **H. F. Richards.** *Physical Review*, New York, N. Y., vol. 22, 1923, pp. 122–23.

9. STATIC ELECTRICITY, **Francis B. Silsbee.** *Circular C438*, National Bureau of Standards, Washington, D. C., June 1942.

Teletypewriter Selective Calling Applied to Automatic Airlines Reservations

R. W. BRATSCHI
NONMEMBER AIEE

W. D. PENNINGTON
NONMEMBER AIEE

A. L. WHITMAN
MEMBER AIEE

A RECENT PAPER has described the *83B1* teletypewriter selective calling system.[1] The UNIVAC airlines reservation system as a special-purpose application of a general-purpose data processor has also been described.[2] The present paper continues the description of the combination of the two, with special emphasis on the alterations, relatively minor, of the standard teletypewriter selective calling system to permit its use as the communications link between the central processing equipment and various remote agent offices to which the public comes to ask questions, reserve seats, or buy tickets.

A second kind of interconnection will be mentioned here but not discussed in detail. For agent sets located near enough to the processor so that it is practical and economical to run numerous wires between them, a different technique, involving a higher speed of operation, is used. Such a situation is encountered where the processor is in the same building with a large reservations bureau, where a number of telephone clerks receive incoming calls from airlines customers.

However, for offices at distances where it is neither practical nor economical to supply a large number of control channels in parallel, one or more single-channel teletypewriter circuits provide the connecting links. The traffic load per station is low enough so that multistation line operation is possible, at a speed of 100 words per minute, with the necessary automatic controls being supplied by the selective calling system.

Basic Operational Features

As a basis for the more detailed description of what the communication equipment has to do, the principles of operation of the system will be briefly reviewed. At a ticket counter, or other location where airlines personnel meet customers, UNIVAC furnishes a device called an agent set, shown in Fig. 1, which serves as a transmitter of inquiries to the processor, and a receiver of answers from the computer. The agent set can also request the sale of seats, upon which the processor corrects its space inventory file and sends a confirmation of the sale to the agent.

In operation, on hearing a question from a customer, a ticket agent selects a projection slide containing a portion of the airlines timetable showing the flight legs in which the customer is interested. The agent inserts this slide in the set, where it is projected on a glass viewing screen. The set automatically identifies the particular timetable slide by photocell reading of edge coding. The agent then sets up his inquiry by pushing buttons opposite the appropriate timetable lines and columns on the screen and by pushing additional buttons to indicate the date, the number of seats, and the type of transaction. The completion of the entry automatically initiates the sequence of operations which result in the setting up of a connection to the processor. Once the connection is established, it is held until the reply comes back and is registered in lights on the agent set.

The function of the selective calling system is essentially to "poll" (search) the agent set locations, and when one is found with an inquiry to send, to establish the connection to the processor and then, after the transaction is completed, to disconnect and continue its search for others with inquiries ready. Actually, the system does not poll each individual agent set, but rather a number of programmer-scanners each of which has subcontrol of up to eight agent sets; see Fig. 2. Once a programmer-scanner gets a chance to use the line it, in turn, offers one chance to each of its satellite agent sets before it disconnects.

Typical System Layout

Fig. 3 is a typical layout of an airlines reservation system, showing the relation of the reservations processor to the local high-speed programmer-scanners and to the remote telegraphic-speed scanners reached over the teletypewriter selective calling system. Fig. 4 shows in more detail the principal elements of the selective calling system at the control station (at the processor location) and at the various outlying stations.

Operation of Selective Calling System

For the polling operation, the transmitter start circuit operates as it does in a teletypewriter message system. Each programmer-scanner location is assigned a unique 2-letter code, known as a transmitter-start code (TSC). The control station sends these codes out in turn, searching for a programmer-scanner which has an inquiry ready to come in.

The entry of an inquiry or transaction into an agent set, as a result of an agent pushing the proper buttons, causes the programmer-scanner to "bid" for a connection to the processor. With a bid thus registered at the time that the station is polled, polling stops, and the inquiry or transaction is handled. On the other hand, if the programmer-scanner has not made a bid when it is polled, the station

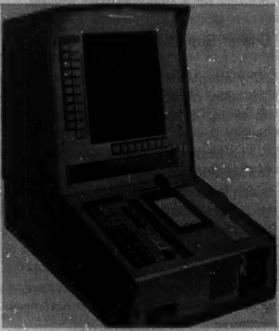


Fig. 1. UNIVAC agent set

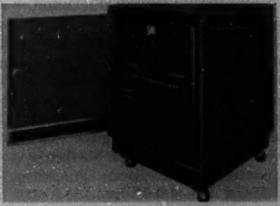


Fig. 2. Telegraphic programmer-scanner

Paper 60-861, recommended by the AIEE Telegraph Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting, Atlantic City, N. J., June 19–24, 1960. Manuscript submitted March 17, 1960; made available for printing April 27, 1960.

R. W. BRATSCHI is with Remington Rand UNIVAC ®, St. Paul, Minn.; W. D. PENNINGTON is with the American Telephone and Telegraph Company, Cincinnati, Ohio; and A. L. WHITMAN is with the Bell Telephone Laboratories, Inc., New York, N. Y.

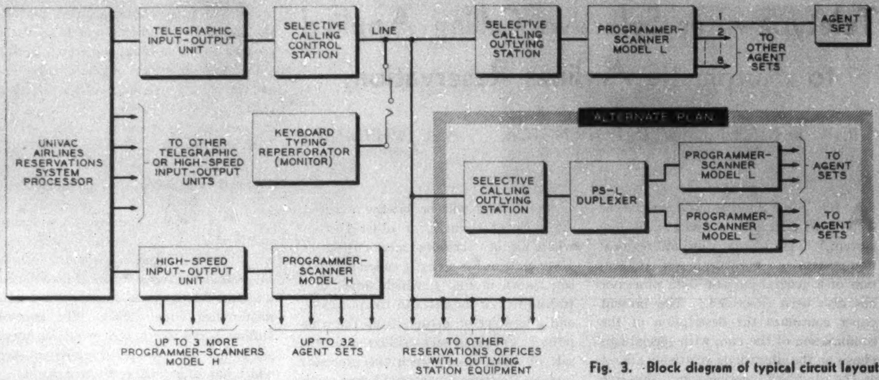

UNIVAC AIRLINES RESERVATIONS SYSTEM PROCESSOR
TELEGRAPHIC INPUT-OUTPUT UNIT
SELECTIVE CALLING CONTROL STATION
LINE
SELECTIVE CALLING OUTLYING STATION
PROGRAMMER-SCANNER MODEL L
1
2
8
AGENT SET
TO OTHER AGENT SETS
TO OTHER TELEGRAPHIC OR HIGH-SPEED INPUT-OUTPUT UNITS
KEYBOARD TYPING REPERFORATOR (MONITOR)
ALTERNATE PLAN
SELECTIVE CALLING OUTLYING STATION
PS-L DUPLEXER
PROGRAMMER-SCANNER MODEL L
TO AGENT SETS
PROGRAMMER-SCANNER MODEL L
TO AGENT SETS
HIGH-SPEED INPUT-OUTPUT UNIT
PROGRAMMER-SCANNER MODEL H
UP TO 3 MORE PROGRAMMER-SCANNERS MODEL H
UP TO 32 AGENT SETS
TO OTHER RESERVATIONS OFFICES WITH OUTLYING STATION EQUIPMENT

Fig. 3. Block diagram of typical circuit layout

returns a no-traffic response, the letter V, and the transmitter start circuit at the control station proceeds to poll the next programmer-scanner in sequence.

Since all inquires are directed toward a single destination, the processor, no receiver selection is required in this application so that no use is made of the call directing codes (CDCs) used for the selective calling of various receiving printers on a teletypewriter system.

Use of Suppression Code Bar

By adaptation, another use is made of this spare function to save wear and tear on some of the teletypewriter mechanisms.

The heart of the selective calling controls is a device called a sequence selector which is, in effect, a model 28 typing unit with the printing and paper-handling features left off. Its main function is to operate and control a "stunt box,"[3] a device which responds to individual teletypewriter characters to operate contacts or to perform desired mechanical operations

Use is made of the stunt-box mechanisms, as described later, to convert serial line signals to the parallel signals needed by the UNIVAC equipment. Operation of these mechanisms is needed at a particular outlying station only when it is communicating with the processor. They are therefore so arranged that their operation is under the control of the "suppression" code bar. (On a typing unit this is known as the "print" code bar; its position determines whether or not printing will take place. On a sequence selector it often is not used.)

However, here this code bar at the active programmer-scanner location is moved out of the suppression position by the sequential combination of the second letter of the transmitter start code, which is unique to the station, followed by the letter E, sent as the start-of-message code by the selected programmer-scanner. Thus at the desired station, the proper mechanisms are conditioned to operate, while at all other stations they remain disabled. On the end-of-transmission signal the operated code bar at the active station goes back to the suppression position.

Line Held for Reply

After a programmer-scanner has been selected, it holds the line until each one of the agent sets tributary to it has had its chance, if wanted, to send an inquiry to the processor and to receive its reply. The programmer-scanner then automatically generates its end-of-transmission signal, the characters HH, which tells the transmitter start circuit to resume its polling search for traffic. Polling continues with the next station in sequence after the one which has just used the circuit.

Codes

All the signals transmitted over the lines are 5-element characters made up of marks and spaces in the same way as is the Baudot code; however, their meanings in the UNIVAC system are different,

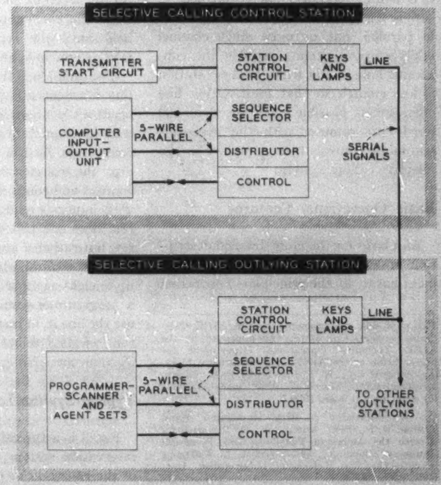

SELECTIVE CALLING CONTROL STATION
TRANSMITTER START CIRCUIT
STATION CONTROL CIRCUIT
KEYS AND LAMPS
LINE
COMPUTER INPUT-OUTPUT UNIT
5-WIRE PARALLEL
SEQUENCE SELECTOR
DISTRIBUTOR
CONTROL
SERIAL SIGNALS
SELECTIVE CALLING OUTLYING STATION
STATION CONTROL CIRCUIT
KEYS AND LAMPS
LINE
PROGRAMMER-SCANNER AND AGENT SETS
5-WIRE PARALLEL
SEQUENCE SELECTOR
DISTRIBUTOR
CONTROL
TO OTHER OUTLYING STATIONS

Fig. 4. Selective calling control equipment

Vol.
Pt. 1
196

and when copy is obtained on a monitoring teletypewriter, it is not readable as normal message text. In all communications between the UNIVAC equipments in the system, only the first four levels of each character are used for the transmission of intelligence; the fifth level is used for an odd parity check, a mark being supplied in this level whenever there would otherwise be an even number of marks in the character. Of the 32 teletypewriter characters, there are 16 having an odd number of marks; these characters and the corresponding UNIVAC codes are listed below:

Teletype Characters	Significance in UNIVAC Message Text
W	0
(Space)	1
Y	2
P	3
U	4
CAR RET	5
B	6
G	7
J	8
M	9
F	'
C	&
LTRS	(
T	"Ignore"*
E	Space
LINE FEED	Minus sign

* T may appear in first four letter positions of reply as well as at end of message.

The significance of the codes is not changed by being preceded by a "figures shift" (FIGS) or "letters shift" (LTRS) but is a function of the position of the character in the message.

Certain of the above codes have control functions in addition to their use within the message text. Each input message to the processor begins with an E, a character not employed in any of the transmitter start codes. Until this code is received, the processing equipment ignores all the start codes and other signals appearing on the line, with the one exception described below. After a message starts, the receiving equipment at the processor continues to accept characters until a T is received, signifying the end of the input message. T is similarly transmitted at the end of the reply message from the processor. (T as any one of the first four letters of the reply has an information function.)

If an error is detected in an input message, the processor sends back a reply notifying the agent of the fault. If it is a parity error, he is generally instructed to re-enter the message to determine whether or not the fault is of a transient nature. If it is a logical error (e.g., a nonexisting date or flight) an appropriate message is transmitted to the agent so that the correction can be made. If an error is detected in the reply message, the programmer-scanner sends a control message (LINE FEED T) to the processing equipment requesting a repeat of the last reply. LINE FEED is the only character accepted by the processor receiving equipment which is not preceded by an E.

The characters having even numbers of marks are used as transmitter start codes and for other control functions performed by the communications equipment, as follows:

HH, *end-of-transmission*

Sent by a programmer-scanner when it is ready to give up the line. Also transmitted as part of FIGS HH LTRS by the selective calling control station to clear the circuit after an idle-line timeout. This code restores the communications equipment to a polling condition.

FIGS, *station cutoff alarm*

Sent by the selective calling control station as part of the code FIGS HH LTRS after an idle-line timeout, this code causes an audible alarm at the outlying station which has control of the line, and an electrical alarm at the corresponding programmer-scanner. For an optional variation, see the further discussion in the section "Alarms and Protective Features."

AA, *alternate arrangement*

Sent from the central processing site, in case of an outage, to instruct all agents to use temporarily other means of communication, such as the telephone. At each outlying station the control signal, through the programmer-scanner, lights a lamp on each agent set.

Serial-to-Parallel Conversion

Present-day teletypewriter transmission is accomplished by means of serial signals, i.e., signals where the successive signal elements of the characters follow each other in time series on the line. For printing and tape-perforating operations, where a received character is represented by five of the serial signal elements, conversion from serial to parallel is effected mechanically by a device known as the processor mechanism. For the UNIVAC computer input, which requires an electrical input of five signal elements in parallel, a new means was needed. Advantage was taken of the flexibility of the stunt box in the sequence selector.

Six stunt-box slots are used, one for each of the five signal elements of the incoming character and the sixth for a timing and control pulse. Each of the five slots is equipped with a function bar so coded that it responds and closes a contact only when the signal element corresponding to its position is a marking element. Thus, when all of the serial elements have been registered through the selector and the associated mechanism, the five contacts are in the marking (closed) or spacing (open) conditions respectively, depending on the signal elements in the five positions. The contact closures are read into the UNIVAC equipment, after which the contacts open and are ready for the next character.

Of the characters used in this system, (except H and A) eight have marks and eight have spaces in each of the five signal element positions. Assuming that the various characters occur with equal frequency, there would be a marking contact operation in each element position for half of the characters sent. Unless otherwise controlled, these contact operations would occur at every outlying station on a line; this would not affect the UNIVAC equipment because the programmer-scanners at the inactive stations would disregard operations which they were not conditioned to receive and understand. It is to save wear and tear on the mechanisms, as described previously, that the operation of these contacts is suppressed at all stations except the one busy with an inquiry.

Parallel-to-Serial Conversion

Standard conventional teletypewriter means were available for performing the parallel-to-serial conversion. A small multicontact cam-operated distributor, economically mounted on the sequence-selector base and driven by the same motor, accepts the parallel signals from the UNIVAC equipment at either end of the circuit and distributes them serially for transmission on the line.

System Performance

An important characteristic of a reservations system is its speed as manifested by system traffic capacity and the message delivery times at the agent sets. For a given system configuration and processor program, limits are set for both of the quantities. The maximum number of transactions the processor can handle in an hour is 3,600 divided by the processing time in seconds, for current systems approximately 5,000 transactions. The minimum delivery time is the sum of the times required for polling transmission, and processing, currently about 3 seconds. The maximum delivery time, as would be experienced if every agent set in the system made an entry at the same instant, is of the order of 1 or 2 minutes. The actual delivery time fluctuates between these limits, with variations in message length

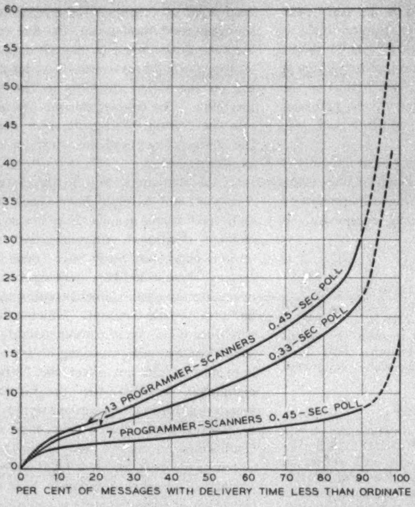


Fig. 5. Computed message delivery times for a particular system

Table I. Types of Messages

	Per Cent of Total Traffic	Number of Characters in a Message	
		Input	Reply
Inquiry	60	12	6
Inventory	30	16	2
Flight status	10	8	3

and in program time, but primarily because the entry is made at random with respect to the time at which that agent set is given a turn to communicate with the processor, resulting in a variable number of other sets which must be processed first. For current systems the actual average response times are generally under 10 seconds.

Prior to putting the first systems into operation, a system simulation was made on a computer at the Bell Laboratories to give a qualitative indication of the sort of performance which might be expected. The method is described in a recent paper.[4] Because of the inevitable uncertainties about the applicability of the assumed parameters to any other system, it should be clear that these results are not of general application. They are thought to be valuable as a general indication of what this kind of system can do, and as a guide to the possible effects of changing certain of the parameters.

Fig. 5 shows the results for three sets of assumed conditions. The delivery time is the time from the moment when the agent has completed the setting up of the message in his set to the moment that the answer is shown on the indicator lights. For reasons associated with the computer simulation program, the curves for delivery times for cumulative percentages above 90 are less reliable; they appear as dashed lines in Fig. 5. The "poll" time is the time per station taken by the selective calling system in searching the stations for traffic. The 0.45-second figure is for the present standard design, using 2-letter calling codes. Where the total number of stations is small enough so that a single letter would be adequate, a modification of the design could be made to give a reduction in delivery times as shown by the curve marked "0.33-second poll."

Even greater benefits, as shown by Fig. 5, could be obtained by reducing the load on the selective calling line by removing some of the stations and putting them on another line. Thus, while with the originally planned arrangement, 90% of the reply messages would be delivered in 30 seconds or less, with the assumed reduced load that time would be reduced to about 8 seconds.

The details of the computation are given in an unpublished memorandum by George J. Smith, formerly of the Bell Laboratories. As background information some of the more significant assumed parameters are extracted here.

The reservations processor was assumed to be handling, in the busy hour, 1,020 transactions over the selective calling circuit and 3,200 transactions locally via the high-speed programmer-scanners; see Fig. 3. There were 44 agent sets connected to 13 outlying low-speed programmer-scanners, seven of these sets at reservations offices and 37 of them at ticket offices. It was assumed that a reservations clerk could enter 40 transactions an hour, and a ticket agent, because of the extra time required in preparing tickets, only 20 transactions an hour. The types of messages were assumed to be as in Table I.

Similar system simulations have been carried on in the St. Paul UNIVAC laboratories, where models of the current systems have been set up in the form of computer programs which can be easily modified to reflect proposed changes in equipment arrangements. Generally the results are summarized in the form of a family of curves showing the average response time as a function of the transaction volume in a peak hour. Each curve in the family represents a particular processor program time. These studies have indicated that changes in program times have results comparable to those produced by the changes in polling time shown in Fig. 5, and that a major factor in agent set response is the employment of an adequate number of separate circuits with the traffic equally distributed among them.

Testing of Stations

The connecting line from each station passes through a telegraph test room, known as the serving test center (STC), in order that the plant maintenance forces may have access to the circuit to locate and isolate troubles which may occur. To avoid unnecessary trips of maintenance men to stations, and to assure that the right man goes, since different groups of men take care of the telephone company equipment and the UNIVAC equipment, it is important to be able to learn as much as possible from the test center as to the nature and location of the trouble.

In ordinary teletypewriter operation, clues as to the nature of trouble can often be obtained by observing the printed copy; since the copy in this case is not English text, it does not give as much information about trouble. Symptoms of trouble are rather the automatic error indications or complete failure to operate.

As one means of helping the isolation of trouble, a second loop, known as the test loop, is provided from the test center to the station. When the station attendant, on request from the test room, operates a specified key, the UNIVAC equipment is cut off from the telephone company equipment, the 5-wire parallel output of the sequence selector is looped back to the 5-wire parallel input of the distributor, and the distributor is arranged to send back on the test loop instead of the regular loop. Thus the test room can send out signals on the regular loop and receive them, for observations of distortion, on the test loop. The test room can thus establish whether there is trouble in the regular loop, or in the *83B1* station equipment. The use of the standard Baudot start-stop code makes possible the use of standard measuring equipment at the test room.

Arrangements can also be made to provide each outlying station with a test slide in the agent set slide cartridge. This is arranged to send a prearranged processor inquiry message, which will elicit a fixed processor reply; the test room connects a monitor printer or reperforator on the circuit and observes the known copy as a clue to the nature of the trouble.

Alarms and Protective Features

The selective calling system is provided with a number of automatic features to safeguard continuity of service, with associated alarms to warn station attendants of trouble.

If during the polling of stations, one does not respond with either a message or the designated no-traffic signal (the letter V), the transmitter start circuit waits about 2 seconds, gives a momentary alarm, and then goes on to poll the other stations. Thus, an interruption from one station does not tie up the system. This momentary alarm repeats on every round of polling so long as the trouble persists. For sure identification, an attendant can operate a key which will stop the polling process the next time the round reaches the offending station.

An open-line condition, occurring during transmission, will cause an alarm. If this should occur during polling, the polling would stop; the continuous search is itself a valuable check on the continuity of the circuit.

If at any time the line remains idle and marking for more than about 10 seconds (subject to change by the user) the transmitter circuit will time out and send the end-of-transmission code automatically. This will clear out any connection which was held up, and alarm the station which caused the holdup. Before the station can make another bid for the circuit, the attendant must push an alarm release button to put out the alarm lamp and remove the lockout condition.

For cases where the attendant has other duties to perform and may be some distance from the *83B* key and lamp panel, an option was added to remove the necessity of manually releasing the alarm. With this option the timeout signal is passed to the programmer-scanner, but the audible and visual alarms on the *83B* equipment do not operate. At the *83B* control station a momentary audible and visual alarm is given. A counter can be provided if the customer wishes to keep a record of the number of time-outs.

If the processor is out of order, an "alternate arrangement" signal (the letters AA) is sent from the processor site to alarm all stations and to tell them that for the time being they must use alternate means of handling their reservations business.

Conclusions

The application of the teletypewriter selective calling system provides a simple and economical communications link for the UNIVAC airlines reservations system. Several airlines are now using the system for a part of their reservations requirements and its use is expected to grow.

References

1. THE 83B1 TELETYPEWRITER SELECTIVE-CALLING SYSTEM, C. W. Smith, C. J. Votaw, A. L. Whitman. *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 78, May 1959, pp. 119–24.

2. THE UNIVAC AIR LINES RESERVATIONS SYSTEM: A SPECIAL-PURPOSE APPLICATION OF A GENERAL-PURPOSE COMPUTER, D. K. Sampson, V. E. Herzfeld, C. W. Fritze. "*Proceedings*, Eastern Joint Computer Conference," *AIEE Special Publication T-114*, July 1959, pp. 152–56.

3. SIMPLIFIED PRINTING TELEGRAPH SWITCHING AND INTEGRATED DATA PROCESSING, J. B. Booth, R. H. Klich. *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 75, Mar. 1956, pp. 11–14.

4. SIMULATION OF DATA-SWITCHING SYSTEMS ON A DIGITAL COMPUTER, F. J. Gross. *Ibid.*, vol. 78, 1959 (Jan. 1960 section), pp. 796–800.

A New Control Terminal for Mobile Telephone Service

L. E. GETGEN
NONMEMBER AIEE

GREG McGIBBON
NONMEMBER AIEE

THE *L-2* control terminal which is described in this paper, is a basic element in a complete "package" radio system. The system design is building block in nature with only a few building blocks required for a complete operating system that is devoid of frequently unnecessary embellishments. However, additional panels are available where more than the basic functions are required.

As the functions of the various panels, the possible system arrangements, and system engineering considerations have been discussed in other literature,[1] they will not be considered here. The circuits used in the control terminal are, however, somewhat unusual and have not been fully described elsewhere. Their operation will be the subject of this paper.

Control Terminal Functions

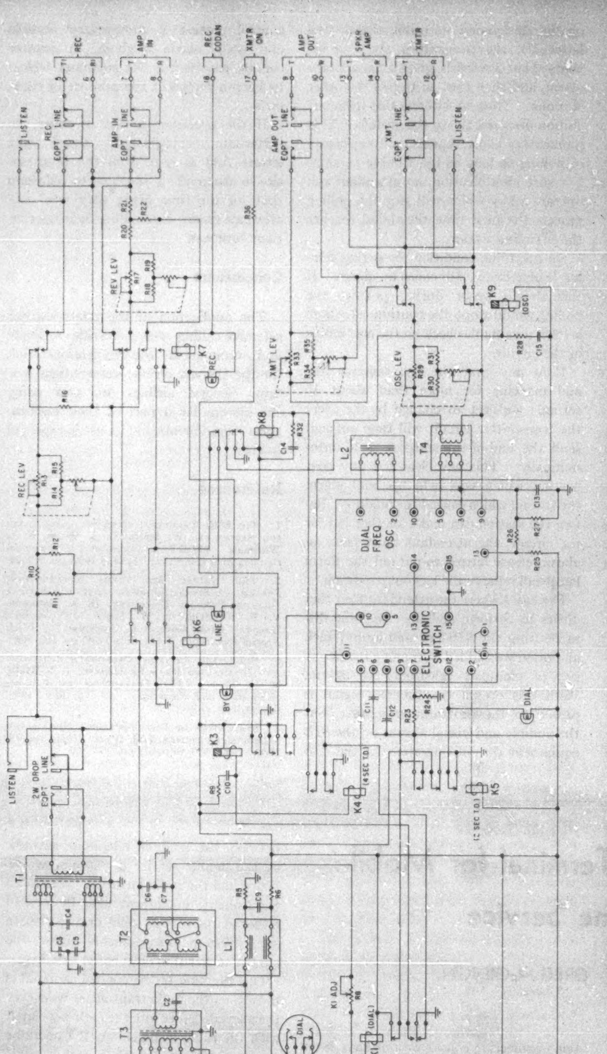
The purpose of the *90101A* control terminal is to provide the interconnection between the landline telephone network at the mobile service operator's switchboard and the radio facilities. The radio transmission equipment may be located either at the central office or at a remote site. System arrangements allow the control terminal to be at either location.

One function of the control terminal is to turn on the radio transmitter whenever the operator plugs into the jack associated with the radio channel. If the operator was initiating a call to a mobile unit, rather than responding to a mobile initiated call, the control unit should also start the 600–1,500-cycle signaling

---

Paper 60-159, recommended by the AIEE Radio Communication Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., January 31–February 5, 1960. Manuscript submitted October 30, 1959; made available for printing March 7, 1960.

L. E. GETGEN and GREG McGIBBON are with the Lenkurt Electric Company, Inc., San Carlos, Calif.


REC
AMP IN
REC CODAN
XMTR ON
AMP OUT
SPKR AMP
XMTR
LISTEN
REC EQPT LINE
AMP IN EQPT LINE
AMP OUT EQPT LINE
XMT EQPT LINE
REV LEV
XMT LEV
OSC LEV
REC LEV
K7 REC
K8
K9 (OSC)
L2
T4
DUAL FREQ OSC
ELECTRONIC SWITCH
K6 LINE
BY
K3
K4 (4 SEC T.D.)
K5 (2 SEC T.D.)
DIAL
T1
T2
T3
L1
TEL SET LEV
K1 ADJ
K1 (DIAL)
K2
L.O
2W DROP EQPT LINE
2W DROP EQPT AS REQD
TEL SET
LOCAL CONTROL NORMAL LOCKOUT
SIC
SID
SIE
(CHASSIS)
SUPV AS REQD
GRD
PWR ALM
-48V DC

oscillator and connect it to the transmitter.

Frequency shift signaling with 600 and 1,500 cps (cycles per second) is the most widely used method of selecting the desired mobile unit in common-carrier mobile service today. In the so-called "manual" systems the dialing is performed by the mobile service operator at some type of manual switchboard. Each dial pulse is represented on the radio channel by a frequency shift. Common practice is for the numeral 1 to be reserved for resetting or clearing the mobile unit selectors. Therefore only combinations of the numbers from 2–10 are used in the mobile subscriber's number.

As mentioned previously, when the mobile service operator in calling a mobile unit, plugs into the radio channel trunk circuit, two things take place. First, the base transmitter is placed on the air. Second, the signaling tone is transmitted. This condition is maintained for 2 seconds and then the control terminal shifts the signaling tone and at the same time provides a "start dial" indication to the operator. The frequency shift, in effect, dials a 1 to reset the selectors in all mobile units. Subsequent dialing by the operator shifts the oscillator in accordance with the dial pulses and thereby, selects the desired mobile unit. Four seconds after the final dial pulse, the control unit automatically dials a 1 to reset all mobile unit selectors and to stop the ringing in the selected mobile unit.

For mobile initiated calls to land subscribers, the base transmitter is turned on when the operator answers but the signaling tones are not used.

Circuit Design

The heart of the *90101A* control terminal is the transistor circuitry shown in Fig. 1 as the electronic switch and the dual frequency oscillator.

Physically, each of these transistor circuit assemblies is a $3^1/_2$- by $5^1/_4$-inch glass-epoxy, printed-circuit card. For ease of maintenance, one transistor type is used throughout.

Fig. 1. Schematic diagram of Lenkurt L-2 control terminal

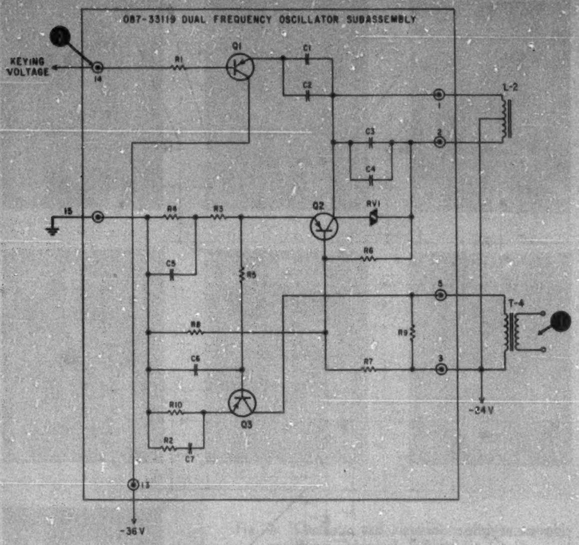

087-33119 DUAL FREQUENCY OSCILLATOR SUBASSEMBLY
KEYING VOLTAGE
14
R1
Q1
C1
C2
1
L-2
C3
2
C4
15
R4
R3
Q2
RV1
R6
C5
R5
5
T-4
R8
R9
C6
R7
3
R10
Q3
-24 V
R2
C7
13
-36 V

Fig. 2. Schematic diagram of dual frequency oscillator

Note: See Fig. 7 for waveforms indicated at circled points 1 and 2

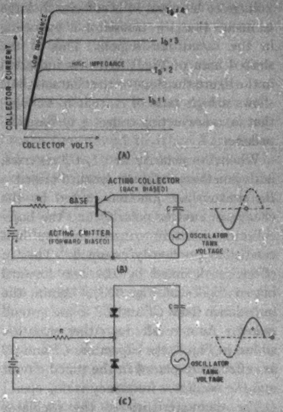

$I_b=4$
$I_b=3$
$I_b=2$
$I_b=1$
LOW IMPEDANCE
HIGH IMPEDANCE
COLLECTOR CURRENT
COLLECTOR VOLTS
(A)
ACTING COLLECTOR (BACK BIASED)
R
BASE
C
ACTING EMITTER (FORWARD BIASED)
OSCILLATOR TANK VOLTAGE
(B)
R
C
OSCILLATOR TANK VOLTAGE
(C)

Fig. 3. Operation of oscillator keying transistor

The oscillator assembly is shown schematically in Fig. 2. The purpose of this unit is to provide 600 or 1,500 cycle signaling tones as directed by the electronic switching assembly.

Transistor *Q2* is a conventional Hartley oscillator with *L2*, *C3*, and *C4* forming the 1,500-cps tuned circuit. Varistor *RV1* provides amplitude limiting. The oscillator output is taken from the emitter and coupled to transistor *Q3* via resistor *R5*. Transistor *Q3* is a common-emitter, class *A* amplifier which provides the signaling tone output through transformer *T4* to the control unit. Resistor *R2* causes sufficient a-c current feedback to maintain the output at approximately 0 dbm (decibels above one milliwatt) with little distortion. Typical waveforms are shown in Fig. 7-1.

Keying of the oscillator is performed by transistor *Q1*. When the *Q1* base is more positive than terminal *13*, *Q1* is biased to cutoff with the collector acting as emitter. In this condition the transistor appears as a very high impedance in series with *C1* and *C2*, and the oscillator tuned circuit consists only of *L2*, *C3*, and *C4*. The resulting oscillator output is 1,500 cps.

When the base is more negative than terminal 13, two conditions are encountered. These depend upon the polarity at *C1* and *C2* in respect to a-c ground. If the *C1*, *C2* point is negative, the emitter base junction is back biased and the collector-base junction is forward biased. This situation is depicted in Fig. 3(B). It will be noted that these bias conditions are correct for transistor action except that in this particular case the collector is being used as the emitter and the emitter is used as the collector. The bias

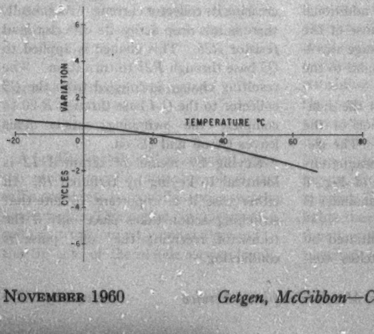


Fig. 4 (left). Temperature characteristics of dual frequency oscillator at 600 cps

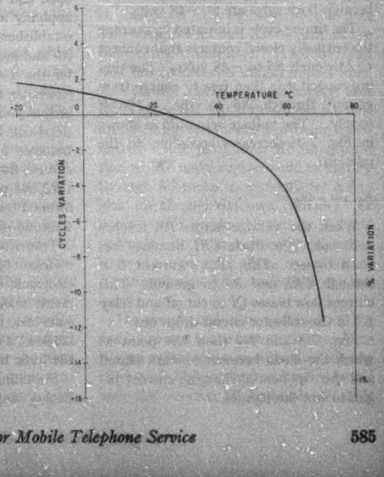


Fig. 5 (right). Temperature characteristics of dual frequency oscillator at 1,500 cps

voltage to terminal 14 is sufficiently large to insure that the transistor is operating in the saturation region. This is the shaded area of Fig. 3(A). As indicated in the figure the slope of the characteristic shows a high ratio of current to voltage; that is, a low acting-collector to base impedance.

When the polarity at *C1*, *C2* reverses, both junctions of *Q1* are forward biased—the emitter-base junction caused by the change in emitter polarity and the base-collector junction caused by the bias conditions previously assumed. This, in effect, is depicted by the two forward biased diodes of Fig. 3(C). Again, the impedance from *C1* and *C2* to a-c ground is low. As a result, for either situation shown in Fig. 3, the capacitors *C1* and *C2* are effectively placed in the tuned circuit and the oscillator frequency is 600 cps.

The characteristics of the oscillator assembly as a function of temperature are shown in Figs. 4 and 5. It will be noted that the frequency variation over the operating temperature range of −20 to +50 C (degrees centigrade) is considerably less than 1%. This is well within the system requirements.

The electronic switching assembly shown in Fig. 6 consists of a 2-second time-delay circuit, a 4-second time-delay circuit, and a bistable switch. To simplify maintenance, a single type of transistor is used in all these circuits.

In the quiescent state, the 4-second-time-delay transistor *Q1*, is biased on through the path from −24 volts, the *Q1* emitter-base junction, resistor *R3* and *R4* to the −48-volt supply. Diode *CR1* is back biased as its anode is tied to the −48-volt supply through normally closed contacts and its cathode is at −25 volts. Time-delay capacitor *C12* has no charge because both sides are at −48 volts.

The timing cycle is initiated by opening the normally closed contacts that connect *C12* through *R5* to −48 volts. The timing capacitor is then free to charge from ground through *R6* to the −48-volt supply. The voltage across *R6* as shown in Fig. 7-9, decreases according to the relation:

$$E_{R-6} = -48\epsilon^{\frac{-T}{RC}}$$

When the voltage across R6 reaches −25 volts, the diode *CR1* becomes forward biased. This allows current flow through *CR1* and *R6* to ground. This current flow biases *Q1* to cut off and relay *K5* in the collector circuit drops out.

Figs. 7-8 and 7-4 show the point at which the diode becomes forward biased and the exponential charging current begins to flow through *R4*.

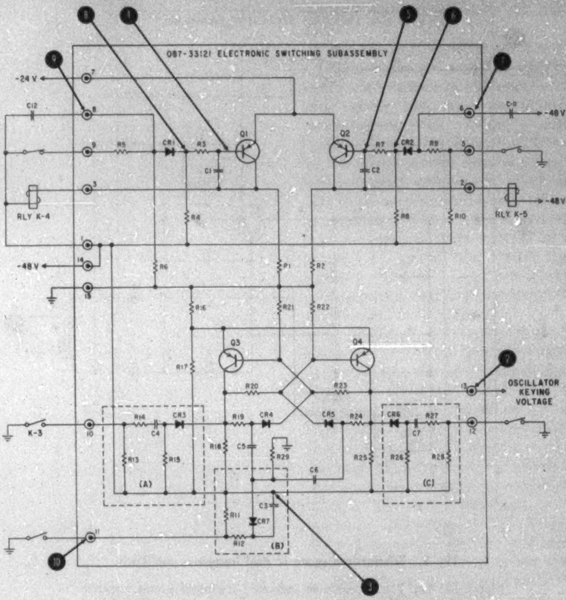


Fig. 6. Schematic diagram of the electronic switching assembly

Note: See Fig. 7 for waveforms indicated at circled points 2–10

The 2-second time delay circuit consisting of transistor *Q2*, and its bias network, is completely analogous to the 4-second delay circuit except for a difference in time constant elements and in the method of starting the timing cycle. In the quiescent state, the switching contacts are open and the timing cycle is started by applying a ground at terminal 5. This establishes a charge path for *C11* through *R9* and subsequent action is the same as for the 4-second delay circuit. Fig. 7-7 shows the voltage across *R9*. At the two second point, a change in the exponential slope can be noted. This is where *CR2* becomes forward biased and additional current flows through *R9* because of the *CR2*, *R8* path. The base voltage waveforms, Figs. 7-5 and 7-6 are similar to the 4-second circuit counterparts.

Transistors *Q3* and *Q4* form the symmetrical bistable switch portion of the electronic switching assembly. The elements making up the pulse forming circuits are those in the boxes of Fig. 6 labeled (*A*), (*B*), (*C*). The remainder is the basic bistable switch.

Switching action may be initiated by closing any one of the switches connected at terminals *10*, *11*, *12*. If switching at terminal *10*, for example, is to be considered note that with *K3* contacts open there is no charge on *C4* as both sides are tied to a common voltage through *R13* and *R15*. Diode *CR3* is back biased in this condition. When *K3* is closed, *C4* charges, principally through *CR3* and *R18*, to form a positive pulse across *R18*. This pulse is coupled via *R20* to the base of *Q4*. This pulse is of the polarity to turn *Q4* off. Therefore, if *Q4* is in the off state when the pulse arrives, no switching action takes place. However, if *Q4* is on, the pulse attempts to turn it off by decreasing its collector current. As a result, there is less drop across its collector load resistor *R25*. This change is applied to *Q3* base through *R23* to turn it on. The resulting change is coupled from the *Q3* collector to the *Q-4* base through *R*-20 to complete the switching action. This leaves *Q4* off and Q3 on.

Keying by means of terminal *12* is identical to keying by terminal *10*. In either case it is important to note that switching action takes place only if the transistor receiving the "off" pulse is conducting.

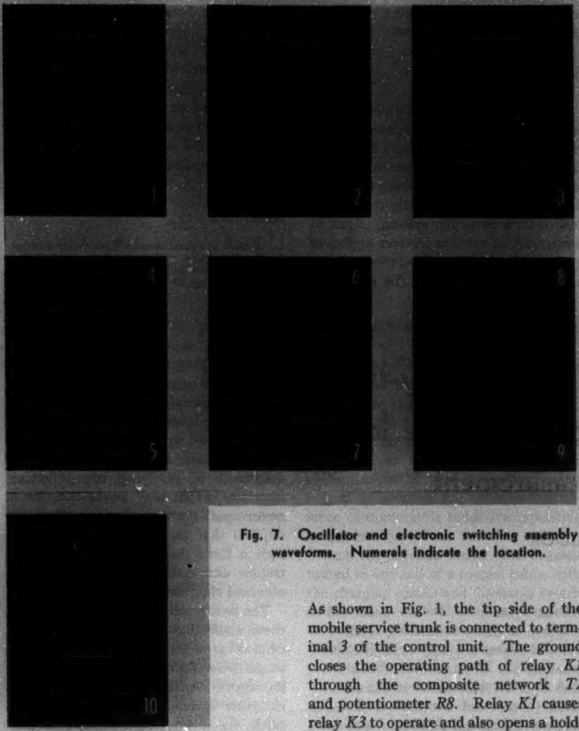


Fig. 7. Oscillator and electronic switching assembly waveforms. Numerals indicate the location.

Keying by terminal *11* is similar to that of the other two terminals with the exception that the trigger pulses are applied to the bases of both transistors. Thus, there is a change of state for each keying pulse regardless of which transistor is on. Figs. 7-10 and 7-3 show dial pulses followed by a 4-second delay pulse at the (B) network input and output.

The purpose of the bistable switch is to provide the appropriate keying pulses to the signaling oscillator. The output is taken from the collector of *Q4* at terminal *13*. Typical waveforms are shown in Fig. 7-2.

Control Terminal Operation

LAND TO MOBILE CALL

A land to mobile call is initiated when the mobile service operator plugs the calling cord into the mobile service switchboard jack. This action, by means of a suitable trunk circuit, applies a ground to the tip side of the mobile service trunk. As shown in Fig. 1, the tip side of the mobile service trunk is connected to terminal *3* of the control unit. The ground closes the operating path of relay *K1* through the composite network *T2* and potentiometer *R8*. Relay *K1* causes relay *K3* to operate and also opens a holding path to the 4-second time-delay circuit (electronic switch terminal *9*). Relay *K3* lights the busy lamp, keys the base transmitter via terminal *17*, operates relay *K9*, and applies ground to terminals *5* and *10* of the electronic switch. Relay *K9* connects the signaling oscillator to the transmitter input terminals.

System operation requires an initial signaling frequency of 600 cps. This condition is assured by the ground applied to switch terminal *10* by relay *K3*. The ground applied to terminal *5* starts the 2-second timing period as previously mentioned. After 2 seconds has elapsed, the 2-second time-delay circuit causes *K5* to release. Relay *K5*, then, removes the remaining −48-volt holding path from electronic switch terminal *9* (one was removed by *K1*) to start the 4-second timing period. It also lights the "dial" lamp at the control terminal and by sending a ground through the composite network *T-2* and the ring side of the line, lights the operator's dial lamp. Another function of *K5* is to apply a ground to terminal *12* of the electronic switching assembly to shift the tone to 1,500 cps. This, in effect, dials a 1 to reset the selectors in all of the mobile units.

Upon receipt of the start dial indication, the mobile service operator can proceed to dial the desired mobile unit. The dialing operation intermittently breaks the ground to relay *K1*. Relay *K1* operates the electronic switch in accordance with the dial pulses via the ground to the electronic switch terminal *11*. The electronic switch, in turn, shifts the dual frequency oscillator.

Each dial pulse, by means of the connection to electronic switch terminal *9*, also resets the 4-second time-delay circuit. Four seconds after the final dial pulse, the 4-second time-delay circuit releases relay *K4*. Relay *K4* applies a ground to the electronic switching assembly terminal *11*, to dial a 1 which stops the ringing in the selected mobile unit and also resets all of the mobile unit selectors. Relay *K4* also de-energizes slow-release relay *K9* which removes the signaling oscillator from the transmitter input.

MOBILE TO LAND CALL

To initiate a call, a mobile subscriber first listens to ascertain that the radio channel is idle and then momentarily pushes his transmit button. Receipt of the r-f (radio-frequency) carrier by the base station receiver causes a ground to be placed on terminal *18* of the control unit. This operates relay *K7*, while the carrier is on, to light the "receive" lamp. Relay *K7* energizes relay *K6* which locks up through its own contacts and contacts of *K8*. In operating, it lights the control unit "line" lamp and sends a ground through the composite network *T2* to light the mobile service operator's line lamp.

When the operator responds, a ground is placed on the tip side of the line. This energizes *K1* and *K3* to turn on the base station transmitter. It will be noted that this condition is similar to that when the operator initiates a call except that *K6* is energized. It is this difference that keeps the signaling tones from being transmitted. The path from *K8* to ground through *K6* make contacts and to battery through *K3* make contacts allows *K8* to lock up through its own contacts. In doing so, *K8* breaks the path to *K9* and prevents it from applying signaling tones to the line. The operation of *K8* also releases *K6* to extinguish the line lamp.

The circuit has been arranged so the line connected to the base station receivers is open except when an r-f carrier is being received. This action is controlled by *K7* and the receiver Codan relays. This prevents the radiation of receiver noise in case of faulty squelch operation during intervals when the base transmitter is on and a mobile unit is not transmitting.

During the periods when the mobile unit is transmitting, *K7* not only connects the receive line to the mobile service operator, but also inserts a reverting level attenuator in the transmit path. This provides the proper level to the base transmitter for the mobile signal. The retransmitted mobile signal is used for mobile to mobile calls and permits other mobile units to monitor both sides of the conversation, as in conventinal party-line service.

Summary

The *L-2* control terminal which requires only $10^1/_2$ inches of rack space can provide all the functions necessary for control of a radio-telephone communication system. It is recognized, however, that additional features are frequently required or desired for a particular system. Most of these, such as satellite receiver appliques for using more than one base receiver with a system; line amplifiers for increasing transmission levels; automatic gain adjusting amplifiers for preventing overmodulation; trunk terminating panels; carrier monitoring devices; and power supplies do not change the requirements of the control terminal. Therefore, they can be added on a modular basis as necessary. The *L-2* control terminal, in each instance, performs the basic control functions.

Reference

1. A New Manual Mobile Telephone System, A. F. Culbertson. *National Convention Record*, Institute of Radio Engineers, New York, N. Y., vol. 7, pt. 8, 1959, pp. 49–58.

Proposed Monocycle-Pulse Very-High-Frequency Radar for Air-Borne Ice and Snow Measurement

JOHN C. COOK
NONMEMBER AIEE

Synoposis: It is suggested that a single electromagnetic pulse having a length in space of the order of a meter can be radiated and received by low-*Q* (broad-banded) antennas. Rough calculations indicate that peak powers of 100 kw probably can be generated with existing electronic components, capable of producing usable signal returns from large semispecular surfaces at distances of the order of miles. The resulting "untuned all-video-radar system" would constitute an electromagnetic analogy to the well-known seismic reflection method of geophysical prospecting, in which a single compressional acoustic pulse (under ideal conditions) is emitted isotropically and reflections are received from successive extensive, flat-lying, reflecting surfaces. A single vhf (very-high-frequency) pulse should combine the advantage of good distance resolution with less tendency than microwaves to be absorbed or scattered by the imperfections in natural materials. Preliminary experiments at 125 mc (megacycles) indicate that interfaces less than 60 cm (centimeters) apart could be so resolved. The development and initial application of the method as a rapid air-borne method of profiling the thicknesses of radio-transparent layered materials, such as floating fresh-water ice (for aircraft landing areas) and the polar continental ice sheets. Potential applications in high-speed extraterrestrial communication may also exist.

Problem

THERE is need for an air-borne means to determine instantaneously and, if possible, continuously, the thickness of floating natural ice masses such as the polar sea ice pack, as possible emergency, scientific, or logistic-support landing fields for aircraft. There appears also to be a need for information concerning the detailed contours of the rock underlying the polar ice caps of Greenland and Antarctica, since these data are now being laboriously garnered by seismic expeditions traveling on the surface. A need has also been expressed for means to determine rapidly the extent and thickness of permafrost in ice-free land areas of the polar regions. In response to these needs, the author proposed in 1956 to develop airborne electromagnetic methods for measuring the thickness of frozen materials, on the assumption that frozen materials, having extremely high electrical resistivity should be more or less transparent to radio waves. A subsequent study, under the sponsorship of the Air Force Cambridge Research Center,[1] has shown that radio waves probably would be severely attenuated in sea ice and permafrost, although useful information might nevertheless be obtained electromagnetically from these types of materials.

With fresh lake ice and continental snow-ice sheets, which are relatively radio-transparent because they have resistivities of the order of 10^6 to 10^7 ohm-meters,[2–4] there still remains the problem of scattering by inhomogeneities within the ice and irregularities upon its surface. Therefore, in accordance with well-known physical principles, electromagnetic energy of the longest possible wavelength should be used to secure deep penetration of these materials. However, the requirement of resolving two reflecting interfaces which may be as little as a meter apart (in the case of floating ice) requires that the radiated energy be in such a form that both adequate penetration and distance resolution can be obtained simultaneously.[5]

The best possible compromise between these conflicting demands would be to employ a radio wave consisting of a single cycle of a vhf carrier as the exploring agent. An analogous single acoustic pulse is employed for probing the earth for plane, acoustically-reflecting interfaces in the seismic-reflection prospecting method.[6] In practice, a single cycle of either acoustic or electromagnetic energy is difficult to generate. However, a series of single pulses (equivalent, by Fourier's principle, to a sum of infinite wavetrains of many frequencies) can readily be generated. After reflection, the waveform received may be different from that transmitted, because of changes in the relative phases and amplitudes of the various frequency components in the equivalent spectrum. From these changes the dispersion, absorption, and layering of the reflecting body and hence its structure, dielectric constant, and re-

---

Paper 60-994, recommended by the AIEE Radio Communication Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Pacific General Meeting, San Diego, Calif., August 8–12, 1960. Manuscript submitted March 21, 1960; made available for printing July 13, 1960.

John C. Cook is with Southwest Research Institute, San Antonio, Tex.

The author wishes to thank The Air Force Cambridge Research Center, in particular Dr. James A. Peoples, Jr., under whose direction this work was supported under Contract no. *AF 19(604)-2137*, for the opportunity to perform this brief preliminary study.

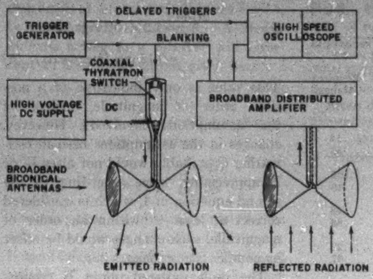


Fig. 1. Hypothetical untuned, "all-video" radar system

sistivity could perhaps be deduced. A narrow search beam would not be required in the case of large, flat-lying reflecting interfaces (any more than it is in seismic prospecting) because rays arriving at angles very far from normal incidence would be reflected away, beyond the region sampled by the receiving antenna. The requirements of the problem and the limitations of vhf electromagnetic pulses therefore seem well suited to one another.

Proposed Radiating System

A single burst or pulse of electromagnetic energy can theoretically be generated by the acceleration of a charge, Q. This amounts to starting or stopping a flow of electric current. The radiation is emitted from the charge itself. In order to generate a single short wavelet, the circuit containing the charge during its acceleration must be kept compact in the intended direction of radiation. A loop radiating edgewise, for example, should be less than half the required pulse "wavelength" in diameter. Directed radiation can be obtained while preserving the coherence and shortness of the emitted pulse only by using a large parabolic reflector, since resonant or multielement antenna structures would radiate a wavetrain when excited with a pulse.

A simple and natural generating antenna for the purpose would consist of two oppositely charged conducting bodies side by side. If these were suddenly connected together, as by a spark or thyratron switch, the flow of charge would be in a straight line and most of the radiation would be emitted sidewise. The charged conductors could be shaped like the halves of a nonresonant, broad-band vhf antenna to encourage radiation of all the frequency components in the spectrum of a short pulse; this is equivalent to prevention of resonant "ringing" and consequent emission of other wavelets subsequent to the first.

One form of such an antenna might be two coaxial opposed cones with their points close together ("biconical" form).[7] The cone angles should be broad and their heights of the order of a quarter of the desired pulse wavelength. A resistive coating might be used to further insure critical damping of the antenna and absence of appreciable inductive-capacitive reverberation.

Such an antenna might also be connected to one end of a coaxial cable, with the charging means and discharge switch at the other end. The discharge-current wave from the switch would be confined within the cable until it emerged into the the exposed conductors of the antenna. Their radiation would occur during a time equal to the sum of the current-wavefront rise time t and the time needed for the current front to spread over the entire antenna. Echoes within the cable could probably be prevented by good matching at the antenna end. Fig. 1 shows a video-radar system with such an antenna.

The switch suggested for such an arrangement is a type *5C22* hydrogen thyratron in series with one or two voltage-multiplying gaps. It would be arranged as the center conductor of a coaxial line which would have tapered matching sleeves at the ends for cable connection. This switch, described by Heard,[8] would be capable of generating discharges of 30 to 50 kv and 500 amperes with a minimum rise time, $t = 1.8$ mμsec (millimicroseconds). Since radiation occurs only during acceleration of the current, which is completed in time t, the corresponding pulse wavelength would be $ct = 54$ cm if only that portion of the circuit adjacent to the switch is considered. Longer radiation pulses would generally be generated by virtue of the fact that additional, finite time is required for the wavefront of current acceleration to be propagated over the surface of the antenna.

Nomenclature

I_θ = radiation intensity in direction θ
Q = radiating charge
t = time, or input pulse duration
θ = elevation angle or phase angle
ϕ = azimuth angle
R = distance from antenna center
c = velocity of light
v = velocity of radiating charge
T = period of radiated pulse fundamental
W = total radiated energy per pulse
q = input charge
C = antenna capacitance
V = supply potential difference
$\lambda = cT$ = pulse wavelength
δ = depth of conduction in antenna skin
n = number of conduction electrons per cm^3
l = distance of charge displacement
P = peak radiated power
G = antenna gain
A = wavefront area
E = electric-wave intensity
r = reflection coefficient of surface
h = height of antenna above surface
p = power density in space
E_s = signal voltage at receiver
F = receiver noise figure
k = Boltzmann's constant
Δf = receiver bandwidth
p_s = received signal power
p_n = receiver noise power
ρ = resistivity of medium
ϵ = dielectric permittivity
α = attenuation coefficient
μ = magnetic susceptibility
$\omega = 2\pi/T$ = angular frequency
Z = receiver input impedance or intrinsic impedance of medium

Estimate of Radiated Peak Power

To calculate the intensity of the electromagnetic pulse emitted by such equipment, an equation derived by Sommerfield for X-ray emission may be used. (See either Smythe's[9] equation 14.12.10; or Stratton's equation 37, p. 475 of reference 10.)

$$I_\theta = \frac{Q^2(dv/dt)\sin^2\theta}{16\pi c^3 R^3 \cos\theta}\left[1 - \frac{1}{[1-(v/c)\cos\theta]^4}\right] \tag{1}$$

is the total pulse energy per-unit area radiated in direction θ, the angle between the directions of v and of R. Gaussian units are used throughout. This must be integrated twice to give the total radiated energy:

$$W = \int_0^\pi \int_0^{2\pi} I_\theta (R\,d\theta)(R \sin\theta\, d\phi) \doteq \frac{Q^2 v}{3c^3}\left[\frac{dv}{dt}\right] \doteq \frac{2Q^2v^2}{3c^3T} \tag{2}$$

In performing the integration, it has been necessary to clear fractions in the bracket, expand top and bottom by binomial theorem, and neglect all terms involving the v/c correction. This can safely be done for the purposes of an order-of-magnitude calculation since v/c is small, according to the following reasoning.

The maximum charge, q, that can be

Table I. Calculated Video-Radar Performance

For Radiated Power $P=10^5$ Watts, Reflection Coefficient $r=0.2$

Wavelength, λ, Meters	Antenna Gain G	Height h, Meters	Signal p, Watts	Signal-to-Noise Ratio, Db	Signal E_s, Volts	Required Gain g, Db
3 meters, 100 mc, $\Delta f=140$ mc, $Z_{in}=$ 3,000 ohms	2	100	0.023	99	5.9	12
		10,000	$(2.3)10^{-6}$	59	0.059	53
	5	100	0.14	107	14	5
		10,000	$(1.4)10^{-5}$	67	0.14	45
0.3 meter, 1,000 mc, $\Delta f=1{,}400$ mc, $Z_{in}=300$ ohms	2	100	$(2.3)10^{-4}$	79	0.18	43
		10,000	$(2.3)10^{-8}$	39	0.0018	83
	12	100	8.3×10^{-3}	85	1.1	27
		10,000	8.3×10^{-7}	45	0.011	67

caused to move through the switch depends upon the antenna capacitance C and the supply potential V, since $q=CV\approx\lambda V/2$ esu (electrostatic units). The movement of q across the gap causes the "compressional" displacement of a much larger charge Q in the surface of the antenna, which does the actual radiating. Q can be estimated from the length of the antenna $\doteq\lambda/2$, its circumference, taken as 1 cm or greater, the "skin depth" $\delta=6.6\sqrt{\lambda/c}$ cm, and the valence-electron charge per cm^3 of copper, n, which is, very roughly:

$$n=\left(\frac{2\text{ electrons}}{\text{atom}}\right)\left(6\times10^{23}\frac{\text{atoms}}{\text{mol}}\right)\times\left(\frac{1\text{ mol}}{63\text{ grams}}\right)\left(\frac{8.9\text{ grams}}{\text{cm}^3}\right)\times(4.8\times10^{-10}\text{ esu})=8.3\times10^{13}\text{ esu/cm}^3$$

$$\text{and } Q=\frac{\delta n\lambda}{2}=6.3\times10^{9}\,\lambda\text{ esu} \quad (3)$$

where mol is the abbreviation for gram molecular weight. The distance "effectively moved by q" (by the electric wave spreading over the antenna surface) is approximately from the midpoint of one side of the antenna to the midpoint of the other, approximately $\lambda/4$. The average distance l, actually moved by Q is much smaller: $l=(q/Q)(\lambda/4)$. The time required is at least the pulse period T, so, assuming $V=30$ kv $=100$ esu, and $T=t+\lambda/4c=2.3\times10^{-9}$ seconds, the mean charge velocity is:

$$v\leqslant\frac{l}{T}=\frac{q\lambda}{4QT}=\frac{\lambda V}{4\delta nT}=2{,}530\text{ cm/second} \quad (4)$$

for the values assumed, whereas $c=3\times10^{10}$ cm/second, so $(v/c)<<1$ and may be neglected in the formula for I_θ.

Taking the value for v estimated in equation 4 and inserting it in approximate equation 2, the average radiated power P during the pulse is:

$$P=W/T=\frac{2Q^2v^2}{3c^3T^2}=1.2\times10^{12}\text{ ergs/second}$$
$$=0.12\text{ mw (megawatt)} \quad (5)$$

This value is only indicative of a possible order of magnitude, because of the assumptions involved. However, changes in the assumptions used in estimating Q probably would not alter W or P appreciably, since according to the second equation in 4, which is considered correct at least to within an order of magnitude, such changes would be offset by compensating changes in v.

The peak radiated power is proportional to v^2, hence to the square of the applied voltage V. It also depends slightly upon the pulse duration T, which is the only other independent variable; the pulse wavelength λ and the antenna dimensions, hence all other quantities, are determined once T has been chosen.

Reception of VHF-Pulse Radiation

Typical broad-band vhf antennas have a "gain" G of the order of 2 to 7, compared to a hypothetical isotropic radiator. The peak radiated power flowing through each square cm at a distance R from a vhf antenna emitting pulses of total peak power P is:

$$\frac{dP}{dA}=GP/4\pi R^2 \quad (6)$$

When the pulse wavefront encounters a smooth plane reflector of reflection coefficient $r=E/E_0=\sqrt{p/p_0}$, the ray paths are folded back and reach the source re-

Table II. Transmission and Reflection Properties of Frozen Materials at 100 Mc*

Layer	Layer Medium	ϵ/ϵ_0	ρ, Ohm-Meters	Attenuation Coefficient α, Meters^{-1}	Working Depth d, Meters	Velocity, Meters/Second	Impedance Z, Ohms	Interface Reflection Coefficient, r	Phase Angle ψ, Degrees	Per Cent Energy Reflected
1	Air	1	∞	0	∞	$(3)10^8$	377			
2	Pure ice	3.2	10^6	1.04×10^{-5}	86,000	$(1.67)10^8$	212	−0.281	180	7.8
3	Pure water	81	10^4	2.1×10^{-3}	420	$(0.33)10^8$	42	−0.67	180	45
1	Air	1	∞	0	∞	$(3)10^8$	377			
2	Snow-firn	1.4	10^6	1.6×10^{-4}	5,700	$(2.52)10^8$	320	−0.082	180	0.62
2′	Glacier ice	3.2	$(5)10^6$	2.1×10^{-5}	43,000	$(1.67)10^8$	212	0 (gradual transition)		
3	Granite	8	3,000	0.022	40	$(1.67)10^8$	$134+j1$	$-0.225+j0.01$	178	5.1
1	Air	1	∞	0	∞	$(3)10^8$	$377+j0$			
2	Cold top sea ice	3.4	200	0.51	1.8	$(1.57)10^8$	$198+j262$	$-0.085+j0.49$	99.8	25
2′	Warm bottom sea ice	4	20	3.6	0.25	$(1.17)10^8$	$101+j65.5$	0 (gradual transition)		
3	Sea water	88	0.3	33.6	0.027	$(0.33)10^8$	$11.6+j10$	$-0.78+j0.03$	177.9	62
1	Air	1	∞	0	∞	$(3)10^8$	$377+j0$			
2	Frozen ground	4	500	0.19	4.7	$(1.48)10^8$	$188+j1$	$-0.33+j0.01$	178.2	11
3	Thawed subground	(30)	10	8.8	0.10		$60+j10.4$	$-0.52+j0.06$	170.7	26

* For normal incidence only.

gion again at a total path distance $R=2h$, twice the flying elevation h. A similar antenna acting as a receiver will intercept an area of wavefront given by:

$$A=G\lambda^2/4\pi \tag{7}$$

The peak power intercepted by the receiving antenna is therefore:

$$P_s=\left(\frac{G\lambda^2}{4\pi}\right)r^2\left(\frac{GP}{4\pi(2h)^2}\right)=\frac{G^2\lambda^2r^2}{64\pi^2h^2}P \tag{8}$$

Assuming that the antenna is matched to the receiver input, one half of this power is available for amplification. The signal voltage E_s at the receiver input is given by:

$$E_s=\sqrt{\frac{p_sZ}{2}} \tag{9}$$

where Z is the input impedance, which is of the order of 3,000 ohms for typical miniature tubes at 100 mc. Suitable 1,000-mc amplifiers are not known to the author, but an input impedance of 300 ohms will be assumed. This formula permits estimation of the required gain, $g=20 \log (25/E_s)$ db (decibels), to drive the display oscilloscope. The voltage required for a 1-cm deflection of typical high-speed oscilloscopes is 25 volts.

The noise power already present at the receiver input is:

$$p_n=FkT(\Delta f)=5(4\times10^{-21})1.4c/\lambda \text{ watts} \tag{10}$$

Here the receiver has been assumed to have a noise figure $F=5$, a value commonly attained in good vhf practice. The bandwidth figure $\Delta f=1.4f=1.4c/\lambda$ results from the requirement of critical damping ($Q=0.76$) to prevent ringing and prolonging of the video pulse into a wave-train in the antenna and receiver. The power signal-to-noise ratio in db at the receiver output is therefore (for λ in meters):

$$10 \log\left(\frac{p_s}{p_n}\right)=10 \log\left(\frac{G^2Pr^2\lambda^3}{3575\pi^2ch^2}10^{-21}\right)$$
$$=10 \log(1.2\times10^{11}\lambda p_s) \tag{11}$$

Table I summarizes some numerical results obtained by applying these formulas to an air-borne ice-surveying situation, for several combinations of extreme values for height h, antenna gain G, and pulse wavelength λ. A peak transmitted power of $P=10^5$ watts and a minimum specular reflection coefficient $r=0.2$ (see Table II) are assumed throughout. It is seen that the signal-to-noise ratios are all ample and, in fact, the transmitted power might be greatly reduced for some combinations of conditions.

Suitable video amplifiers for a receiver[11] are available only for the longer wavelengths. Spencer-Kennedy Laboratories' model *202D* chain amplifier has a pass band of from 1 kc to 210 mc, a gain of 20 db and an output voltage of 4 volts. An additional special stage of two would be desirable to drive the oscilloscope. This arrangement should be usable up to flying heights of 1,000 meters. If two chain amplifiers were cascaded, operation at 100 mc up to $h=10{,}000$ meters should be feasible.[12]

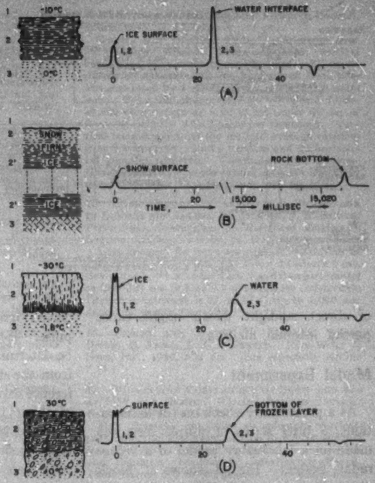


Fig. 2. Hypothetical appearance of reflected video pulse sequences from various terrains

A—Lake ice layer 200-cm thick
B—Continental icecaps 1,500 meters thick, over granite
C—Average sea ice 200-cm thick, in winter
D—Frozen ground 200-cm thick, over wet gravel

Application to Ice Measurement

Table II presents values of the vhf radio-reflection coefficients of various types of interfaces between natural layered materials common in the Arctic and Antarctic. The resistivities ρ and dielectric constants ϵ/ϵ_0 used were obtained from the literature[13–15] or experiment.[1] The attenuation coefficient α, velocity v, and useful working depth d in each material were calculated by assuming a maximum permissible absorptive attenuation of $E_0/E=6$ in the theoretical formulas, using mks (meter-kilogram-second) units:[16]

$$E/E_0=e^{-\alpha z},\ d=(\text{logn } 6)/2\alpha \tag{12}$$

$$\alpha=\omega\sqrt{\frac{\mu\epsilon}{2}\left[\sqrt{1+\frac{1}{\epsilon^2\rho^2\omega^2}}-1\right]} \tag{13}$$

$$v_p=\frac{1}{\sqrt{\frac{\mu\epsilon}{2}\left[\sqrt{1+\frac{1}{\epsilon^2\rho^2\omega^2}}+1\right]}} \tag{14}$$

The complex reflection coefficients r were calculated (except those given by Saxton[17]) by means of the Fresnel formula:[18]

$$r_{12}=\frac{Z_2-Z_1}{Z_2+Z_1} \tag{15}$$

where Z is the intrinsic impedance of a medium, given in mks units by the equation:

$$Z=\sqrt{\frac{\mu}{\epsilon(1+1/j\omega\rho\epsilon)}} \tag{16}$$

Fig. 2 summarizes the hypothetical effects of various natural reflectors upon a video pulse according to the predictions of Table II. At vhf, for relatively low values of resistivity ρ, as in sea ice, the medium would become dispersive (higher-frequency components traveling faster) and also would act as a low-pass absorptive wavefilter, according to equations 13 and 14. The air–ice interface would reflect the high-frequency components better, according to equations 15 and 16; see also reference 19. An attempt has been made to depict these effects qualitatively in Fig. 2. Resistivity data, necessary to determine the wave velocity in the ice hence its corrected thickness, could probably be calculated from spectrometer analyses of the transmitted and received waveforms.

Best results in measuring thickness would be expected with fresh-water lake ice and continental glaciers. According to Stern[2] any measurements involving radio waves should be made when the ice is cold and dry, preferably in the absence

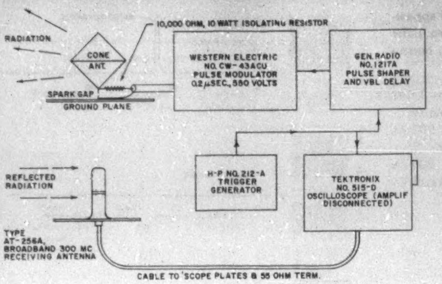

RADIATION
CONE ANT.
10,000 OHM, 10 WATT ISOLATING RESISTOR
SPARK GAP
GROUND PLANE
WESTERN ELECTRIC NO. CW-43ACU PULSE MODULATOR 0.2 μSEC., 550 VOLTS
GEN. RADIO NO. 1217A PULSE SHAPER AND VBL DELAY
REFLECTED RADIATION
H-P NO. 212-A TRIGGER GENERATOR
TEKTRONIX NO. 515-D OSCILLOSCOPE (AMPLIF DISCONNECTED)
TYPE AT-256A, BROADBAND 300 MC RECEIVING ANTENNA
CABLE TO 'SCOPE PLATES & 55 OHM TERM.

Fig. 3. Experimental video-radar test setup

of direct sunlight which might cause partial internal melting.

Model Experiment

As a preliminary check on the assumptions, a brief series of simple tests was made on a laboratory model of a video-radar system. Fig. 3 shows a block diagram of the equipment used.

The radiating equipment consisted of a broad-band cone-and-plane antenna and a pulse-sharpening spark gap. Pulses of direct current at 550 volts, 0.2-μsec long, were supplied by a conventional radar modulator. A series-isolating resistor was used to make the antenna characteristics independent of the modulator and power cable. The ideas for this video source were developed after reading Potok's article.[20]

The receiving system consisted of a wide-band aircraft antenna connected directly to the deflection plates of a fast-sweep oscilloscope through 6 feet of low-loss 55-ohm coaxial cable. None of the available amplifiers, including that in the oscilloscope, had enough bandwidth to pass the expected video pulses. Echoes in the antenna cable were combated by means of a 55-ohm carbon-film terminating resistor at the oscilloscope plates.

The oscilloscope was triggered either before the modulator, as shown, or afterward by the initial transmitting antenna transient, picked up by induction in the trigger cables. Either trigger could be chosen by adjustment of the trigger threshold bias control of the oscilloscope. Pretriggering displayed the entire received waveform, but with so much jitter (0.3 μsec, arising at the spark) that it could not be intensified enough for analysis. Post-triggering stopped all jitter, but failed to display the early portion of the strong wave group leaking into the receiver directly from the transmitter. Available antenna cables provided insufficient time-delay of the received signal to permit viewing the initial pulse. Hence, its waveform could not be determined, although it was estimated from the distances involved in the interference experiments to be a damped train of about 13 125-mc waves, as shown in Fig. 4. This demonstrates that waves of much shorter duration than the trigger pulse can be generated by an antenna spark gap (or a thyratron, probably). Because of the pressure of other work, no serious attempt was made to "clean up" reradiation from nearby objects or to increase the antenna damping to determine whether the wavetrain could be shortened thereby.

It has been suggested by J. D. King, Southwest Research Institute, that such a damped wavetrain could be radiated as a single pulse by employing an alternative technique: the wavetrain would be fed by cables differing in length by λ to two different, equal antennas placed close together. Attenuation in the two cables would be adjusted so that the first and subsequent cycles radiated from one antenna would cancel the second and subsequent cycles from the other, leaving its uncanceled first cycle as the only energy radiated. Presumably the same arrangenent could be used as a nonresonant receiving antenna.

Because of the lack of an amplifier, the radiation could not be seen on the oscilloscope for antenna separations greater than 17 meters. Hence, attempts to bounce reflections from large, fixed flat surfaces (e.g., metal buildings) failed because sufficient range could not be used to separate the reflection from the tail of the direct leakage signal.

However, a movable corner reflector 1 meter square could be identified on the oscilloscope by motion of the interference "kink" produced in the tail of the leakage wavetrain, Fig. 4. A movement by only ±30 cm in range could be detected. It was estimated that, if a suitable amplifier had been available, and if the radiated wavetrain had been properly shortened, fixed reflectors only 60 cm apart could be resolved with this system. Such high resolution has never been achieved as yet, with a pulsed-radar system in the vhf portion of the r-f (radio frequency) spectrum.

Other Applications

It should be pointed out that applications of vhf video radar are limited because of the difficulty of obtaining narrow antenna beam width at such frequencies. However, this is not a serious impediment for specular reflection from large flat surfaces such as natural bodies of ice. Vhf antenna beams of small angular width could be obtained in a monocycle system only by the use of large parabolic reflectors, since multi-element antennas would lengthen the pulse into a wavetrain. Such reflectors probably must either be mounted on the ground or must float weightlessly in extraterrestrial space. Applications to communications in space may arise. It is surmised that there will be circumstances in which attenuation by internal scattering is a limiting factor. Two examples might be: 1. probing the atmospheres of the giant planets by radar, and 2. high-speed telegraphic communication through vast distances filled with rarified, inhomogeneous matter. Under such conditions, the basic idea here proposed might have value: that of combining the required long radio wave and the desired short signal duration in single electromagnetic pulses.

Much sharper distance resolution in microwave radar could be achieved if desired by the monocycle-pulse technique, when suitable broad-banded r-f amplifiers become available. However, the finite angular beam width achievable limits lateral resolution seriously, unless several receiving antennas and interference techniques are used to increase it.

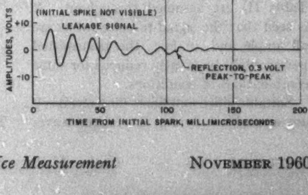


Fig. 4. Observed radiation waveform with reflection

Therefore, resolution in range of a few centimeters might not be worthwhile at present.

References

1. RF Electrical Properties of Salty Ice and Frozen Earth, **J. C. Cook.** *Journal of Geophysical Research*, Washington, D. C., vol. 65, 1960, pp. 1767–71.
Laboratory measurements at 100 mc suggest that r-f resistivities of new sea ice are probably lower than d-c resistivities reported by Dichtell and Lundquist by a large factor. Dielectric constant was about 3.5 at all temperatures.

2. On a Method and Some Results of Electrodynamic Measurement of Glacier Ice Thickness (in German), **W. Stern.** *Zeitschrift für Gletscherkunde und Glasialgeologie*, Innsbruck, Germany, vol. 18, 1930, pp. 24–42.
A Ph.D. Dissertation project. Theory is developed for the use of radio waves reflected from a glacier bottom, either by tuning through the interference pattern of direct and reflected waves between a separated receiver and transmitter, or by measuring the capacitive loading of a transmitting antenna resulting from the supposed high-permittivity bottom plane. Experiments on the latter method gave reasonable and repeatable relative results at 1 mc on ice 10 to 40 meters thick. Drilling was not done to confirm results, unfortunately. Antenna position upon or above snow not critical. Repeatable measurements possible only before sunrise, because of internal melting of ice.

3. Electrical Resistivity of Snow and Ice, **Irwin Roman.** *Bulletin no. 23*, International Association of Hydrology, Riga, Latvia, 1938, pp. 483–91.
Tested the 4-electrode Wenner resistivity method for measuring depth of mountain snowfields. Iron stakes used with spacings up to 5 meters, input 800 volts and 100 microamperes, measured potentials 2 to 100 volts. Surface snow and granite beneath had resistivities of 1 megohm-meter, and the intervening 2 meters of snow had 2 megohm-meters. Method apparently was successful.

4. On the Determination of Glacier Thickness by the Electric Resistivity Method (in Russian), **I. Mikhailov.** *Isvestiya Akademii Nauk SSS*, Moscow, USSR, Geographical and Geophysical Series, 1939, pp. 417–20 (English abstract).
The Wenner method was tried on two mountain glaciers in summer with electrode spacings up to 100 meters, d-c supply at 200 volts and 0.1–0.3 ma, and potential readings up to 10 volts. The upper 40 meters of ice had a resistivity averaging 3×10^6 ohm-meters, while in the lower 60 meters the resistivity rose toward 3×10^7 ohm-meters.

5. Depth Measurements in the Seward Ice Field by Sonic Echo Ranging, **T. D. Northwood, F. W. Simpson.** *Annual Report*, National Research Council of Canada, Ottawa, Ont., Canada, 1948.
Acoustic coupling to glacier firn proved very difficult. Simple tests with radar gave numerous complex echoes; suggests a low-frequency radar method. Authors believe that equipment development far from a test site would be slow and uncertain.

6. Exploration Geophysics (book), **J. J. Jakosky.** Trija, Los Angeles, Calif., 1950.

7. Antennas: Theory and Practice (book), **S. A. Schelkunoff, H. T. Friis.** John Wiley & Sons, Inc., New York, N. Y., 1952, 638 pp.

8. 20-Kilovolt Delta-Function Generator, **H. G. Heard.** *Review of Scientific Instruments*, New York, N. Y., vol. 25, no. 5, 1954, pp. 454–58.
Describes a coaxial switch employing hydrogen thyratrons which can generate 500-ampere pulses with a rise time as short as 0.6 mμsec.

9. Static and Dynamic Electricity (book), **W. R. Smythe.** McGraw-Hill Book Company, Inc., New York, N. Y., 1939, 560 pp.

10. Electromagnetic Theory (book), **J. A. Stratton.** McGraw-Hill Book Company, Inc., 1941.

11. Distributed Amplifier Covers 10 to 360 Mc, **H. Scharfman.** *Electronics*, New York, N. Y., vol. 25, no. 7, 1952, pp. 113–16.

12. Millimicrosecond Oscillography, **Y. P. Yu, et al.** *Ibid.*, vol. 24, no. 7, 1951, pp. 106–11.
Technique for amplifying and displaying pulses 0.5 to 12 mμsecs long.

13. Properties of Ordinary Water-Substance (book), **N. E. Dorsey.** Reinhold Publishing Corporation, New York, N. Y., 1940, 673 pp.
Concerned chiefly with pure ice. Dielectric constants of 4.3 to 2.2 are observed and calculated for temperatures of −1 C (degrees centigrade) to −70 C for 60 kc and up. Resistivity is 33,000 ohm-meters at −1 C to 300,000 ohm-meters at −40 C for 60 kc for distilled water of 5,000 ohm-meters resistivity. Resistivity is essentially frequency-independent above 60 kc.

14. An Investigation into the Physical and Electrical Characteristics of Sea Ice, **W. J. Dichtel, G. A. Lundwuist.** *Publication no. 122, SIPRE document #503*, National Research Council, Washington, D. C., 1951.
Careful and valuable winter experiments near Pt. Barrow, Alaska. Thin, scattered, vertical brine-filled veins and ribbons observed. Electrical resistivity of horizontal ice bars sawed from floes was measured by 4-electrode d-c method. Results were: 200–2,000 ohm-meters at −7 C and 2,000–30,000 ohm-meters at −24 C for upper 60 cm of floes, of salinity 4 to 7 parts per thousand. Vertical bars sawed from the ice had average resistivities 0.1 as great. Forty vertical and 90 horizontal samples reported. Concluded that conventional resistivity spreads cannot measure ice thickness; suggested high-frequency a-c methods.

15. High Frequency Absorption Phenomena in Liquids and Solids, **W. Jackson, J. A. Saxton.** *Proceedings*, Institution of Electrical Engineers, London, England, vol. 96, pt. III, 1949, pp. 77–80.
Loss tangent for pure ice reaches maximum of 2.39 at 40 kc and −5 C. Is below 0.1 above 1 mc. Sea-water loss varies between 3 and 6 nepers/meter at 300–500 mc (0.6 to 0.3 ohm-meters resistivity). Dielectric constant is 81 at 17 C up to 1,000 mc, dropping slowly thereafter.

16. Fields and Waves in Modern Radio (book), **S. Ramo, J. R. Whinnery.** John Wiley & Sons, Inc., 1944, 502 pp. (See especially Article *7.15*.)

17. Reflection Coefficient of Snow and Ice at VHF, **J. A. Saxton.** *Wireless Engineer*, London, England, 1950, pp. 17–25.
Uses Fresnel's equations to calculate vertical-path reflection coefficient of 0.56 to 0.9 for floating ice at 300 mc to 3,000 mc, depending on wavelength and thickness (because of interference effects). Phase change is 0 to 360 degrees. Calculates fresh-ice absorption at 0.02 db/meter for 10-mc to 10,000-mc waves at −12 C to 0 C, dielectric constant of 3.05 over same range. Gives complex permittivities of sea, snow, fresh ice, thawed soil, and curves of reflection versus frequency and incident angle for several combinations.

18. Dielectrics and Waves (book), **A. R. Von Hippel.** John Wiley & Sons, Inc., 1954, 284 pp.

19. Transfer of Transient Electromagnetic Surface Waves into a Lossy Medium, **J. Keilson, R. W. Row.** *Journal of Applied Physics*, New York, N. Y., vol. 30, 1959, pp. 1595–98.

20. Researches into Spark Generation of Microwaves, **M. H. N. Potok.** *Proceedings*, Institution of Electrical Engineers, vol. 103, pt. B, no. 12, 1956, pp. 781–87.
Describes a technique of generating short wave-trains with a pulse of slower rise time.

◆

Discussion

Frank R. Dickey, Jr. (General Electric Company, Syracuse, N. Y.): Dr. Cook's very interesting paper presents convincing arguments for believing that a wide-band vhf radar will be useful for exploration of ice layers. It certainly is to be hoped that further experiments will be made soon to prove whether or not this is indeed the case.

The method of generating wide-band radio waves by means of a spark gap or gas discharge is one in which I happen to have had a little experience. It undoubtedly is the simplest method known for generating the high-power short-duration pulses which are required in the proposed application. Dr. Cook's estimate of the power output from a spark radiator represents reasonable numerical value, although some of the steps taken in his calculation seem to require further justification. In regard to estimating the power output from such a device, I would suggest the expedient of calculating the electrostatic energy in the field surrounding the radiating elements immediately before breakdown of the gap and then of assuming that all of this energy is radiated at radio frequencies and in the lowest frequency mode of damped oscillation. Naturally, this procedure gives only an upper limit, since some of the energy is used in ionizing the gap and in other losses. However, experiments which I made, reported in reference 1, showed it to give at least the right order of magnitude in the case of very small mercury spheres discharging to a ground plane, in which case the efficiency would be expected to be no higher and probably lower than in the vhf case. At any rate, Dr. Cook's estimate of the power output is probably conservative.

Reference

1. The Production of Millimeter Waves by Spark Discharges, **F. R. Dickey, Jr.** *Doctoral Thesis*, Harvard University, Cambridge, Mass., 1951; also *Technical Report 123*, Cruft Laboratory, Harvard University, July 10, 1951.

Edward M. Little (U. S. Navy Electronics Laboratory, San Diego, Calif.): This looks like a promising method for measuring ice thickness from the air. However, at 125 mc, the accuracy is only ±30 cm (±1 foot), so the operation is marginal at 100 mc. At 1,000 mc, 10 times the resolving power, a 1,000-mc video amplifier will have to be available; none exist now.

This method should be usable on first-year sea ice after late spring when a large part of the brine has drained out but before interior melting is severe; and on old, essentially fresh pack ice, especially in the winter when internal and surface melting is at a minimum.

At Shippegan, New Brunswick, I obtained for the top layer of the ice 30 to 50 ohm-meters in February at an air temperature of −7 C and 100–500 ohm-meters in April at an air temperature of +5 C. These values are lower than those of the author's.

The working depths in fresh ice that Dr. Cook tabulates, if obtainable in practice, would make possible continuous thickness measurements over any glacier or ice cap and would be an extraordinary development for glaciologists who now depend on tedious surface seismic measurements.

John C. Cook: The discussions submitted are most encouraging and tend to support my belief that the proposed technique is feasible and potentially valuable. Unfortunately, in the 3 years since the initial studies reported here were completed, funds to develop and test the idea have not been found.

The method suggested by Mr. Dickey for

estimating the radiated power from the stored electrostatic energy is much more direct, less awkward, and intuitively better than the method used in my paper. The stored energy W and the power P available during a discharge time T can readily be estimated from the antenna capacitance C and potential difference V just before discharge: $W=(1/2)CV^2$ and $P=W/T$.

Inserting reasonable values compatible with the assumptions used in the paper, $V=100$ esu, $T=2.3\times10^{-9}$ seconds, and $C=14$ esu (calculated by assuming an antenna consisting of two spheres 12 cm in radius, 36 cm apart center-to-center), the maximum possible radiated energy and power are found to be: $W=7\times10^4$ ergs and $P=3$ mw. It appears that my earlier estimate of 0.12 mw is indeed conservative.

Dr. Little's values for the resistivity of sea ice are roughly comparable to the values of 20 to 200 ohm-meters given in my Table II. As shown in reference 1 of the paper, variations of sea-ice resistivity with salinity (i.e., history), specimen orientation, test frequency, and temperature extend through several orders of magnitude. The values used in Table II represent my best guesses for new sea ice in midwinter (air temperature −25 C) as measured at vhf with the electric field perpendicular to the brine tubules. They may, again, be conservative when applied to mature sea ices (which predominate in some parts of the arctic ocean), since Dr. Little's measured values of resistivity are, to me, unexpectedly high for the air temperatures he mentions. Either direct video-radar tests, or a thorough study of r-f electrical properties should be made on representative arctic sea ices. Such experiments would establish whether the video-radar method, in addition to its potentialities for land-ice exploration, might also be used to measure the thickness of sea ice from aircraft in flight.

Two-Wire Operation of N Carrier Systems

R. W. DEAN
ASSOCIATE MEMBER AIEE

E. M. KARKAR
NONMEMBER AIEE

THE SAN FRANCISCO BAY area, like many other metropolitan areas, has had a tremendous population increase in a very short period of time. More people mean more industry and more homes. The demand for telephone circuits to give these customers the service they desire has created a cable and conduit problem. Short-haul carrier systems have made great strides toward the relief of this problem, but even these have been only a partial answer. The purpose of this paper is to describe an experiment made by The Pacific Telephone and Telegraph Company for further utilization of existing cable plant to meet this rapidly increasing demand for service.

The *N1* carrier system was designed to provide 12 two-way conversations on two nonloaded cable pairs or a quad within the same cable sheath, using separate pairs and a different frequency band for each direction of transmission; one direction to use the low-frequency band 44 to 140 kc, the other direction to use the high-frequency band 164 to 260 kc. The *N1* carrier system has been described in a previous paper.[1] Open-wire carrier systems have for many years used filters to operate on an equivalent 4-wire basis over a single pair of wires. By applying these techniques to the *N* carrier system it is possible to operate them in cables in the same manner as other systems operate on open-wire pairs. For lack of a more descriptive term, the systems operated in this manner have been designated "2-wire *N* carriers," and will be referred to as such in this paper.

General

The cables crossing the San Francisco Bay were selected for 2-wire *N* carrier-system trials because of the paired and/or quadded cables that were available. The Pacific Telephone and Telegraph Company installed four of these systems using Lynch Communication Systems, Inc., *F-1154* filters. The information received from testing these systems can be used as a guide for engineering of similar systems. Fig. 1(A) shows a schematic layout of the two systems installed between San Francisco and Oakland. Fig. 1(B) shows a schematic of the two systems installed between San Francisco and San Rafael. The systems between San Francisco and San Rafael will be the ones more fully described in this paper because they are more complex.

The first two experimental systems were installed between San Francisco Juniper and Oakland Main offices, with repeaters at San Francisco toll office and Yerba Buena Island repeater station. These systems were operated 4-wire in 19-gage quadded toll cable between San Francisco Juniper and San Francisco toll offices. From San Francisco Toll to Yerba Buena Island and on to Oakland Main the systems were operated 2-wire in 19-gage paired submarine cable.

After these systems were put into operation and thoroughly tested with satisfactory results, they were turned over to regular traffic use for a trial period of 6 months. The following five different types of traffic circuits were assigned to these two systems: two full-period leases, four foreign-exchange circuits, nine crossbar tandem to crossbar tendem, four crossbar tandem to *4M* switching machine, three *3CL* switchboard to *4M* switching machine. During this trial period no circuit outages or system trouble was experienced.

The second two systems, with which this paper is most concerned, were installed between the San Francisco main and the San Rafael office. These systems were operated 2-wire from San Francisco Main to San Rafael in 19-gage quadded cable, with a system working on each side of a quad. There are two repeaters between these offices as shown in Fig. 1(B). The one nearest San Francisco Main, Point Cavallo, is an auxiliary repeater and normally receives its power from San Francisco. The one nearest San Rafael, Alto Junction, is a main-power point. The section between San Francisco Main and Point Cavallo is 6.85 miles in length, and is partly submarine cable. The section between Point Cavallo and Alto Junction is 6.67 miles and is buried cable. The section between Alto Junction to San Rafael is also buried cable and is 7.20 miles in length.

The conversion to 2-wire from 4-wire on these systems and the testing procedures were as follows. Both systems were lined up on a 4-wire basis. Noise, crosstalk, and channel response tests were then made on both systems. The first system was then converted to 2-wire operation in the section between San Rafael and Alto Junction repeater. All

---

Paper 60-998, recommended by the AIEE Wire Communication Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Pacific General Meeting, San Diego, Calif., August 8–12, 1960. Manuscript submitted May 10, 1960; made available for printing June 17, 1960.

R. W. Dean is with the Pacific Telephone and Telegraph Company, San Francisco, Calif., and E. M. Karkar is with Lynch Communications Systems, Inc., San Francisco, Calif.

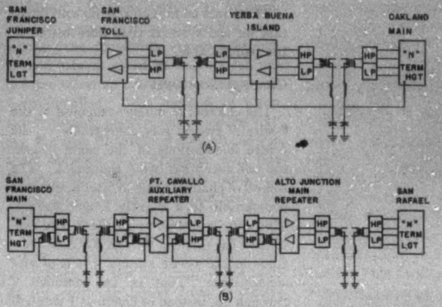


Fig. 1. Schematic layouts of 2-wire systems; see Fig. 4 for equipment details at repeaters

A—Between Oakland and San Francisco Juniper
B—Between San Francisco Main and San Rafael

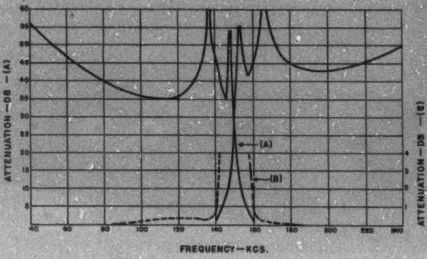


Fig. 3. Response curve of F-1154 high-pass low-pass filter

noise, crosstalk, and channel response tests were then repeated. This procedure was repeated one section at a time until the system was operating 2-wire from San Rafael to San Francisco. The second system was then converted to 2-wire in the section betwen San Rafael and Alto Junction on the pair left idle when the first system was converted to 2-wire. The system continued to operate 4-wire on its own quad between Alto Junction and San Francisco. All noise, crosstalk and channel response tests were then repeated on both systems. This procedure was repeated for each section until both systems were operating 2-wire, one on each side of the same quad from San Francisco to San Rafael.

Filter Design

The low-pass high-pass directional filters were designed by Lynch Communication Systems, Inc. These provide a minimum loop rejection of 60 db (decibels), pass all frequencies below 140 kc on the low-pass side and pass all frequencies over 160 kc on the high-pass side. The schematics of the filters are shown in Fig. 2. The balanced arrangement contains four coils and eight capacitors for the low-pass side, while the high-pass contains three coils and ten capacitors. The impedances in and out are 135 ohms balanced. The frequency response of the filters is shown in Fig. 3. The passband losses for the two channels adjacent to the filter crossover does not vary more than 1/2 db at the corners. The attenuation at the crossover point at 150 kc is a minimum of 30 db to ensure the 60-db minimum-loop loss requirement. The reflection coefficient in both pass bands at no point exceeds 10%.

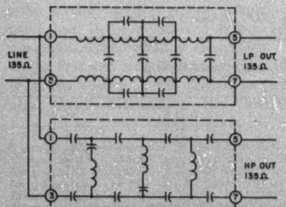


Fig. 2. Schematic of 152-kc Lynch type of F-1154 high-pass low-pass filter

The high-pass and low-pass elements of the filter are separately packaged in a hermetically sealed container measuring $3^1/_4$ by $3^3/_8$ inches by $4^1/_8$ inches high. Two complete filters (four containers) required for one repeater are mounted on a $3^1/_2$ by 19-inch mounting plate, which is arranged to include the battery-feed isolation filters for each side of the repeater. Fig. 4 shows a schematic of the repeater filter and network panel and Figs. 5(A) and (B) show the filter panel. The center container holds the battery-feed isolation filters which are exposed in the front view, Fig. 5(B).

Filter Installation

The filters used for the experimental systems were installed for convenience of cutover from 4-wire to 2-wire and not for maintenance or flexibility. At the terminals the filters were installed in vacant bay space close to the terminals. At the repeater points they were installed close to the frame for convenience of swinging jumper wire.

In a permanent 2-wire installation the filters should be installed near the frame or cross-connect terminals, with both equipment and line side of the filters appearing on frame or terminals, whichever is used. This should be done at both the terminals and repeaters to allow complete flexibility for reassigning filter and associated repeaters for routine or emergency maintenance, or for reassignment of equipment to other systems. Fig. 4 shows a schematic diagram of a repeater, its filters, and building-out equipment.

Transmission Consideration

NOISE

Table I shows a set of noise measurements taken on one of the tested systems. These measurements were taken between the hours of 11 p.m. and 7 a.m. in order to have a minimum amount of disturbance from office dial noise. All noise measurements were made with a *2B* noise-measuring set calibrated 144 and read *F1A* weighted. The noise measurements shown in Table I were made at the +10 point on the channel units (E_1, E_2 jacks), but they have been corrected and are all in decibels adjusted (dba).

If a 2-wire *N* carrier system were to be installed in a route where no power feed was required to auxiliary repeaters, it would still be necessary to use the trans-

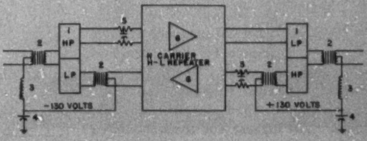


Fig. 4. Detailed schematic of repeater showing filters and power feed

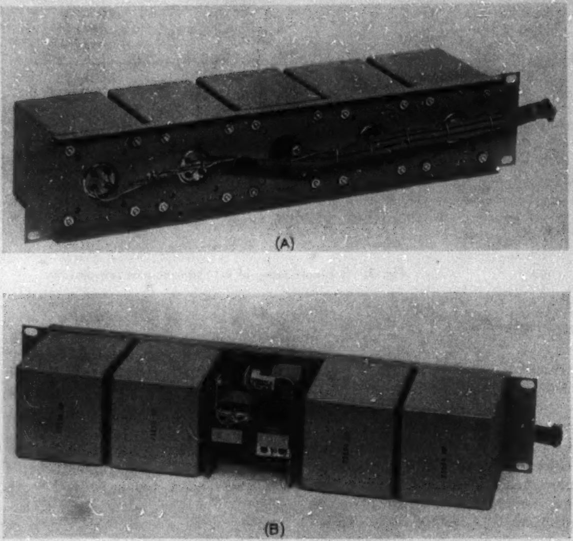


Fig. 5. **Filter mounted on plate**

A—Back view of filters showing wiring
B—Cover removed to show power-feed filters

former with a center tap to ground for suppression of longitudinal noises on the line side of the filters; see Fig. 1(B), San Rafael to Alto Junction section. This was found to be necessary in both experimental installations.

As can be seen in Table I, noise begins to increase noticeably as the third section is converted to 2-wire. Calculations based on these measurements shown in Table I indicate a limit of about five repeater sections should be set on a 2-wire system having repeater sections of between 6 and 7 miles in length. The amount of noise in a different 2-wire *N* system of three repeater sections could be less than the one covered here if the repeater sections were shorter. This is true because the artificial lines and/or span pads used to build out a repeater section are on the 4-wire side of the filters and therefore would add more attenuation in the loop around a repeater. This can be seen in Fig. 4.

CROSSTALK

The two systems covered in this paper were checked for both near-end and far-end crosstalk. The only measurable near-end crosstalk that was found showed up as intrachannel side tones. The highest side tones measured were still down to about −50 dbm (db referred to 1 milliwatt). Most channels were in the order of −60 to 65 dbm. These side tones also appeared in the experimental systems in paired cable, but were not of any consequence.

There was a small amount of far-end crosstalk that could be measured with the measuring equipment available. This only appeared when the systems were operating 2-wire for two sections and 2-wire for three sections. If the far-end crosstalk had been intelligible, it could have been reduced to unintelligible babble by switching systems from one quad to another at repeater points.

Table I. **Noise Measurements in DBA Corrected to 0 Point**

Channels Measured	4-Wire, 3 Sections	2-Wire, 1 Section; 4-Wire, 2 Sections	2-Wire, 2 Sections; 4-Wire, 1 Section	2-Wire, 3 Sections
1	10.1–11.1	11.7–13.7	11.6–12.6	15–16
2	16.6–17.6	12.7–13.7	12.1–14.1	16.1–17.1
2	15.8–17.8	11.6–12.6	10.6–12.6	16.7–17.7
4	12.4–13.4	14.6–16.6	14.6	18–20
5	9.2–10.2	14.4–16.4	13.2–14.2	14.2–15.2
6	10.2–11.2	13.5–15.5	11.3–12.3	14.3–16.3
7	10.4	10.5	11.3–12.3	16.3–18.3
8	11.9–12.8	15.4	11.7–12.7	15.7–17.7
9	8.4– 9.4	8.7	11.4–13.4	12.3
10	7.5– 8.5	7.7	8.7	10.7–12.7
11	9.8–10.8	12.9–14.9	12.8–14.8	14.2–16.2
12	15.0–16.0	12.1–14.1	13.1–14.1	13.2–16.2

There was no measurable far-end crosstalk on the systems that were tested in the paired cable.

CHANNEL RESPONSE

The channel frequency response of the 2-wire *N* system was comparable to the 4-wire installations of the same length, the one exception being channel 1. The roll-off of channel 1 due to filter action was comparable to a slightly longer system. This can be overcome in the 2-wire systems in the same manner as it is now overcome in the 4-wire system if it is of any consequence. That is, use channels 2 through 13, instead of channels 1 through 12.

In both experimental installations it was considered unnecessary to use channels 2 through 13 as channel 1 roll-off was only about 1 db more in the 2,800 to 3,000-cycle range when comparing 2-wire to 4-wire. There was little or no change in the other channels all the way across the frequency band.

Power

The power feeding of auxiliary repeaters presents a problem in a 2-wire system. This problem, however, only exists if there is more than one auxiliary repeater between power points. The principal use of 2-wire *N* carrier will be in metropolitan areas. This means repeaters can be located in central offices with an occasional auxiliary between mains.

In both experimental installations there was only one auxiliary repeater between power points. In the case of the Oakland to San Francisco Juniper systems San Francisco Toll fed negative 130 volts to Yerba Buena Island repeater station on the first system, while Oakland fed positive 130 volts. This arrangement uses a ground return. However, by having San Francisco Toll feed positive 130 volts on the second system and Oakland feed negative 130 volts, the ground currents were equalized. If this method is followed with all 2-wire systems through a cross section, the maximum current through ground return would be one system unbalance when an odd number of systems were installed. This method of alternately feeding positive and negative voltage was also the method used on the systems between San Francisco and San Rafael; San Francisco Main being one power plant and Alto Junction being the other.

At an auxiliary repeater that has both 2-wire and 4-wire systems, the repeater mountings will not be strapped the same as for receiving power from a remote supply. If at any time there should be a need to transfer equipment from a 2-wire to a 4-wire or vice versa, the strapping of the mountings must be changed.

Line-Up and Maintenance

The 2-wire *N* carrier systems require no special equipment for line-up or maintenance. In lining up these systems the standard procedures should be followed. This is possible because this arrangement of 2-wire operation is actually a change in the facility over which the system is operating, and not a change in the system itself. This facility change does means, however, a change in power-feed arrangements and must be remembered for maintenance or reassignment; see the section on power.

Conclusions

Based on the data received from the experimental systems, it has been demonstrated that *N1* carrier systems can now be operated 2-wire for short distances, using available equipment, without exceeding the transmission limits of the systems.

Reference

1. A New Telephone Carrier System for Medium Haul Circuits, R. S. Caruthers, H. R. Huntley, W. E. Kahl, Ludwig Pederson. *Electrical Engineering*, vol. 70, no. 8, Aug. 1951, pp. 692–97.

Establishment of a 450-Mc Vehicular Radio System in the Los Angeles Basin

G. APPLEGARTH
NONMEMBER AIEE

IN ORDER to supply additional mobile-radio-telephone service to the public in the greater Los Angeles area, The Pacific Telephone and Telegraph Company applied for four 450-mc (megacycle) radio channels in June 1956. At that time there were only six channels available for the domestic public mobile-telephone service (MTS) in the 450–460-mc common-carrier radio-frequency (r-f) band.

We have had considerable experience in mobile radiotelephony, particularly in the 460-mc citizen mobile-frequency band. Therefore, certain aspects of transmission and equipment operations on these frequencies were familiar. However, a few unique conditions concerning marriage and compatibility of equipment between radio station and the general telephone plant had to be considered before an integrated system could be established.

The first and most important concern of a new system is the desired service area. After determining the desired coverage, as specified by the Commercial Department, a detailed study of the topography of the greater Los Angeles area and the surrounding territory was made. From our engineering experience with the existing 460-mc private mobile systems and from Kenneth Bullington's radio propagation studies, we decided to establish base transmitters at our Mt. Wilson toll office.[1]

The communication industry considers the Mt. Wilson location a utopia for broadcasting r-f energy. The mountain top is a virtual forest of antenna structures and antennas placed by frequency-modulated (FM) and television broadcasters. Common-carrier and mobile communicators, and miscellaneous common carriers, have also contributed to the array of transmitting antennas. Figs. 1 through 3 are views of antennas and antenna systems on Mt. Wilson.

Mt. Wilson, with an elevation of 5,700 feet above mean sea level, is located 4 miles northeast of the city of Pasadena

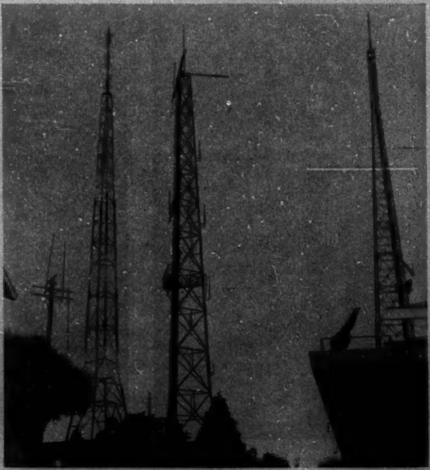
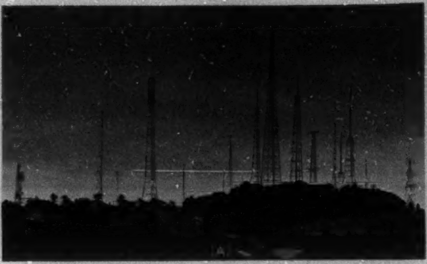


Fig. 1. Center tower, Pacific Telephone and Telegraph Company 450-mc common-carrier transmitting antennas

---

Paper 60-995, recommended by the AIEE Radio Communication Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Pacific General Meeting, San Diego, Calif., August 8–12, 1960. Manuscript submitted March 30, 1960; made available for printing May 26, 1960.

G. Applegarth is with The Pacific Telephone and Telegraph Company, Los Angeles, Calif.



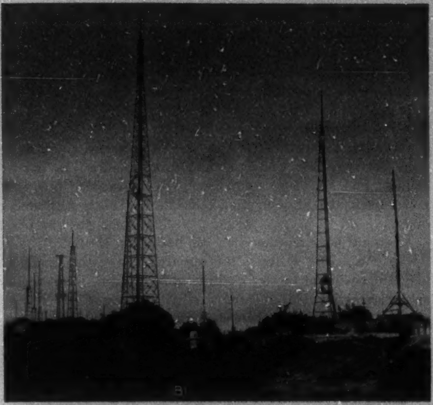


Fig. 2. Tower with three crossarms, center right in A, center left in B, Pacific Telephone Company 450-mc antenna system

A—Mt. Wilson, looking east from Columbia Broadcasting System television location

B—Mt. Wilson, looking west

Calif. In determining the average terrain in accordance with Rule *21.115* of the Federal Communications Commission's (FCC) rules, an elevation of 2,898 feet above mean sea level is derived. Figs. 4(A)–(H) show profiles of the topography taken at 45-degree increments from true north. Rule *21.115* requires profile information between the 2-mile and the 10-mile points on these radials, to determine the average terrain figure. In our particular case the general service area does not conform to this treatment and we are serving mobile subscribers 20 to 40 miles air-line distances from Mt. Wilson. The terrain of the general service area of the Los Angeles basin falls between 100 and 1,000 feet above sea level. This situation gives the Mt. Wilson site an advantage of approximately 5,000 feet above the topography in about 65% of the general service area. Another interesting fact is that the mountain top is in a line of sight from about 75% of the general service area. Fig. 5 illustrates this favorable condition.

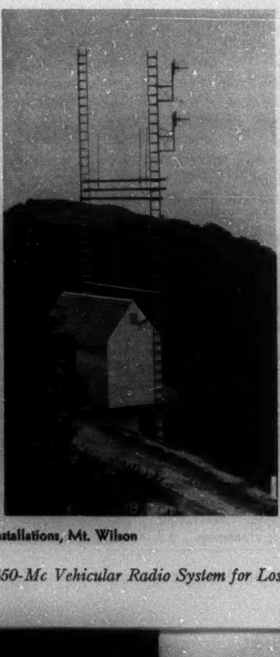
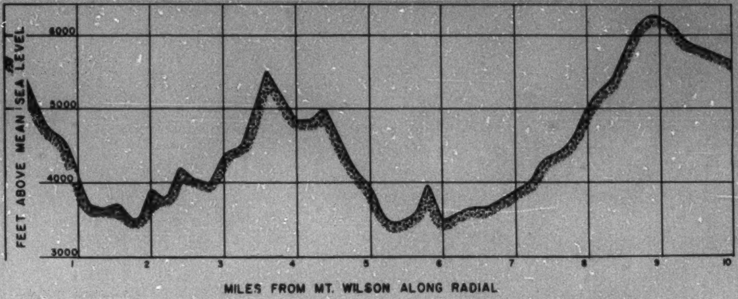


Fig. 3. Private mobile installations, Mt. Wilson

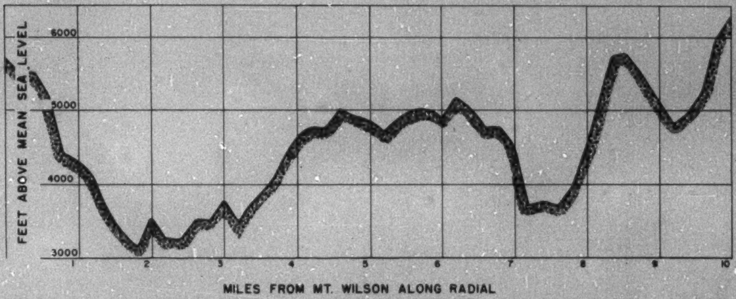
Our company maintained private mobile 460-mc base stations at the Mt. Wilson location for several years prior to the contemplated 450-mc common-carrier offering. From time to time interference problems plagued our private mobile units at this site. However, the nature of the operation permits them to be resolved quite reasonably by moving the carrier frequency a few kilocycles. In the case of the common-carrier offering, interference problems were not as easily resolved. We were necessarily restricted by the fixed assigned frequencies of these channels. Also, new stations at Mt. Wilson were going on the air at a rapid rate. The intermodulation products, direct interference, and off-channel discriminator pulling could not be accurately determined in advance of their contemplated operation. High-intensity r-f fields from high powered r-f transmitters gave many interference problems to the communication receivers when maximum sensitivity was expected. In view of these conditions, it was decided to place only the transmitters on Mt. Wilson and to locate the receivers in various areas throughout the Greater Los Angeles basin.

The topography of the Los Angeles basin created a problem in the determination of receiving sites which would permit a balance between talk-out and talkback. After considerable map study of the area, it was decided to establish four receiver sites to cover the proposed service area. Fig. 5 shows the four locations: 1. Verdugo

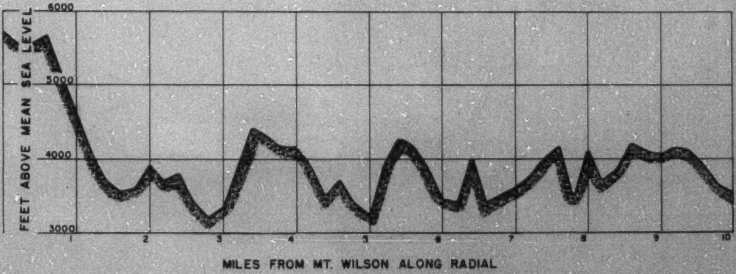
Fig. 4. Profiles of topography, data based on US Geodetic Survey (USGS) series of California topograpic maps



A—Profile along radial bearing 0 degrees true north from Mt. Wilson, USGS1/62500 series, Pasadena and Acton

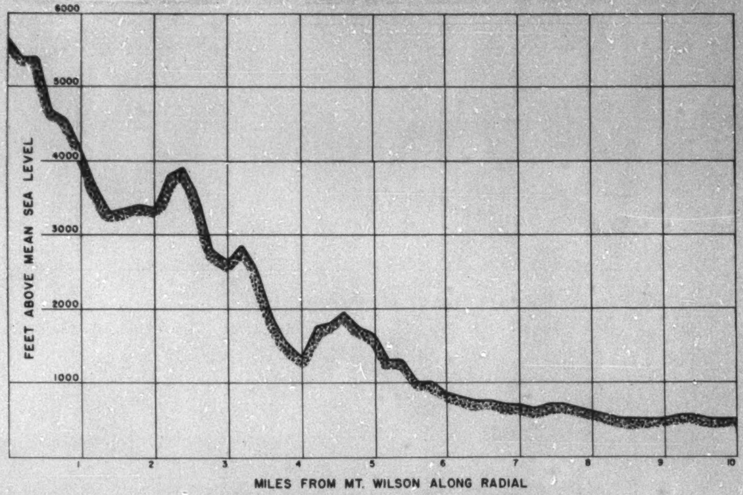


B—Profile along radial bearing 45 degrees true north from Mt. Wilson, USGS 1/62500 series, Pasadena, Acton and Rock Creek

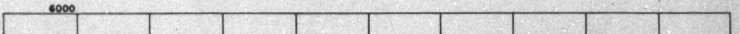


C—Profile along radial bearing 90 degrees from Mt. Wilson, USGS 1/24000 series, Mt. Wilson and Azusa

Fig. 4 (continued).

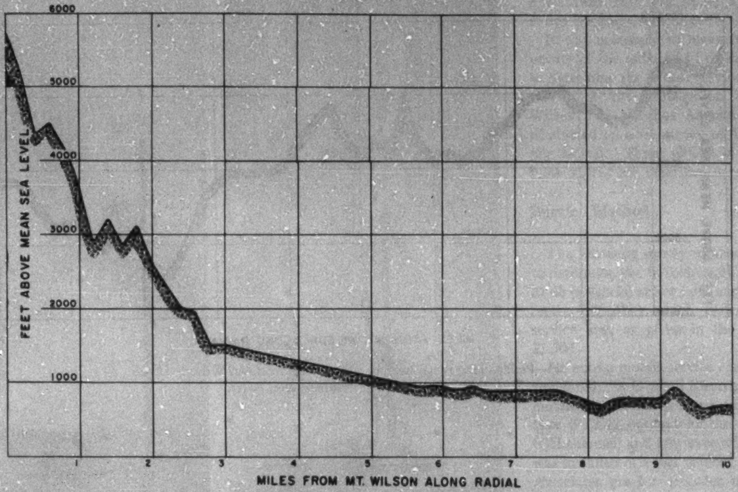


D—Profile along radial bearing 135 degrees from Mt. Wilson, USGS 1/24000 series, Mt. Wilson, Azusa, and Baldwin Park

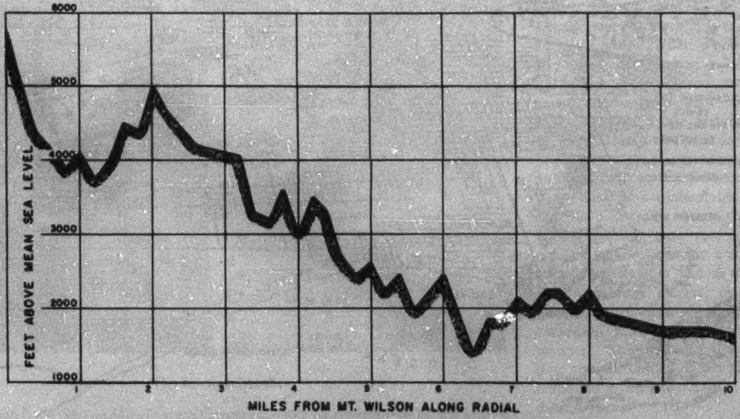


E—Profile along radial bearing 180 degrees from Mt. Wilson, USGS 1/24000 series, Mt. Wilson and San Gabriel

Fig. 4 (continued).

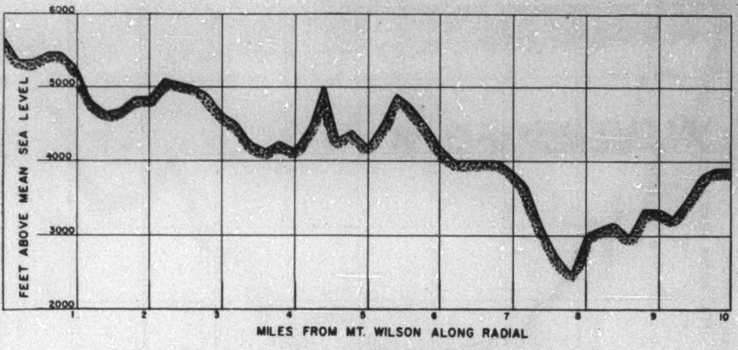


F—Profile along radial bearing 225 degrees from Mt. Wilson, USGS 1/24000 series, Mt. Wilson, Pasadena, and Los Angeles



G—Profile along radial bearing 270 degrees from Mt. Wilson, USGS 1/24000 series, Mt. Wilson and Pasadena

Fig. 4 (continued).



H—Profile along radial bearing 315 degrees from Mt. Wilson; USGS 1/62500 series, Pasadena and Acton

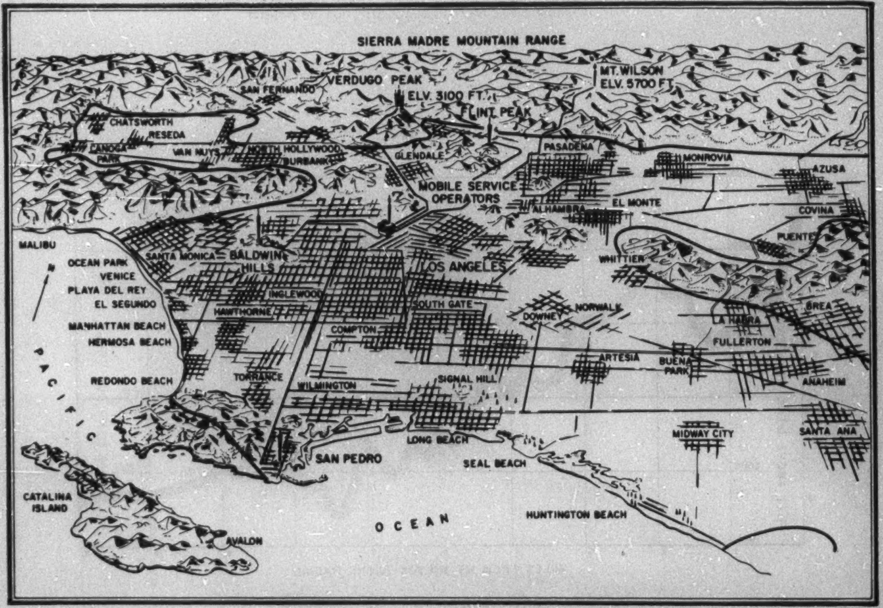

SIERRA MADRE MOUNTAIN RANGE
SAN FERNANDO
VERDUGO PEAK
ELV. 3100 FT.
MT. WILSON
ELV. 5700 FT.
FLINT PEAK
CHATSWORTH
RESEDA
CANOGA PARK
VAN NUYS
NORTH HOLLYWOOD
BURBANK
GLENDALE
PASADENA
MONROVIA
AZUSA
MOBILE SERVICE OPERATORS
ALHAMBRA
EL MONTE
COVINA
PUENTE
MALIBU
SANTA MONICA
BALDWIN HILLS
LOS ANGELES
WHITTIER
OCEAN PARK
VENICE
PLAYA DEL REY
EL SEGUNDO
INGLEWOOD
HAWTHORNE
SOUTH GATE
DOWNEY
NORWALK
BREA
LA HABRA
MANHATTAN BEACH
HERMOSA BEACH
COMPTON
FULLERTON
ARTESIA
BUENA PARK
REDONDO BEACH
TORRANCE
WILMINGTON
SIGNAL HILL
ANAHEIM
PACIFIC
LONG BEACH
MIDWAY CITY
SANTA ANA
SAN PEDRO
SEAL BEACH
CATALINA ISLAND
OCEAN
HUNTINGTON BEACH
AVALON
N

Fig. 5. Artist's concept of 450-mc common-carrier system in Greater Los Angeles basin

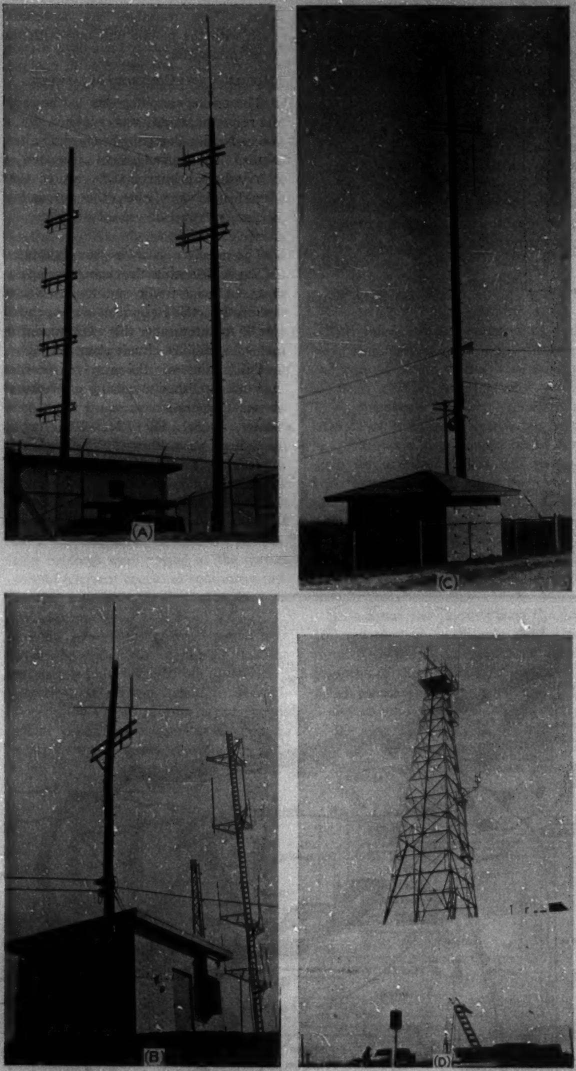


Fig. 6. Receiver sites

A—Verdugo Peak; 450-mc Mark Products Company C-7455 7.2-db gain antenna, top of pole in foreground

B—Flint Peak; 450-mc Mark Products C-7455 7.2-db gain antenna on telephone pole

C—Baldwin Hills; 450-mc Mark Products C-7455 7.2-db gain and Communication Products Company, Inc., 201–509 5.8-db gain antenna, crossarm right

D—San Pedro Central Office; 450-mc Mark Products C-10455 10-db gain antenna, top right of tower

Peak, Fig. 6(A); 2. Flint Peak, Fig. 6(B); 3. Baldwin Hills, Fig. 6(C); and 4. San Pedro Central Office, Fig. 6(D).

It was necessary to measure field intensity of the radiated r-f energy in order to determine the actual service area. A survey of the propagation from Mt. Wilson over the Los Angeles area was conducted on a frequency of 454.750 mc per second. Three phases of the complete survey are described here.

Survey Method

The following survey method was used to determine the field-strength contour of 37 db (decibels) above 1 μv (microvolt) per meter (37 dbu) which defines reliable service area as given in the FCC rule *21.504*.

An actual mobile-service condition was simulated. A General Electric Company mobile transmitter-receiver combination, type *WTE42*, option *0* for the 454.750-mc (QH channel) and 1/4-wave whip antenna was installed in a test vehicle. A Varian Associates graphic recorder model *G-10* was used to measure the grid current of first limiter–last i-f (intermediate-frequency) stages of the receiver. Fig. 7 shows the equipment arrangement. The Varian graphic tape recorder was calibrated to record 37 dbu at a center-scale point on the chart paper. To determine the 37-dbu field strength contour it is necessary first to change limiter-last i-f current recordings to μv input to the receiver, then μv input to μv per meter. The necessary conversion is shown in the Appendix where 37 dbu is approximately 3-μv input to the receiver.

The vehicle was driven over the required service area with concentrated attention given to the periphery and to the internal shadowed areas as determined from several years experience providing mobile service in the Los Angeles area. These data were accumulated by coding each field-strength observation on the Varian graphic tape recorder with a number. Simultaneously, a voice recording of this number and a street intersection, mileage marker, or prominent terrain feature was made which was to be associated later with this observation number. This was also tied to a time base so all recorded data could be synchronized in order to obtain an intelligent data analysis.

Control Measures

Because of different sources of variable performance, a method of control is required in order to minimize spurious

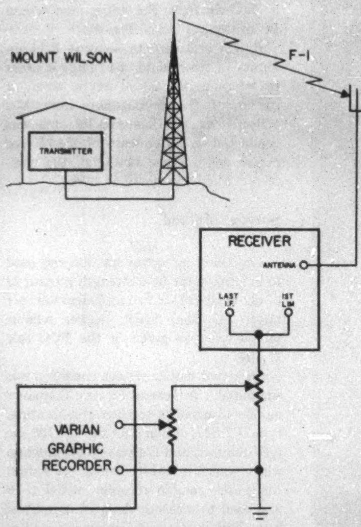

MOUNT WILSON
F-1
TRANSMITTER
RECEIVER
ANTENNA
LAST I.F.
1ST LIM
VARIAN GRAPHIC RECORDER

Fig. 7 (left). Survey test-equipment arrangement

data. These variables are associated both with the mobile unit and with the base transmitter.

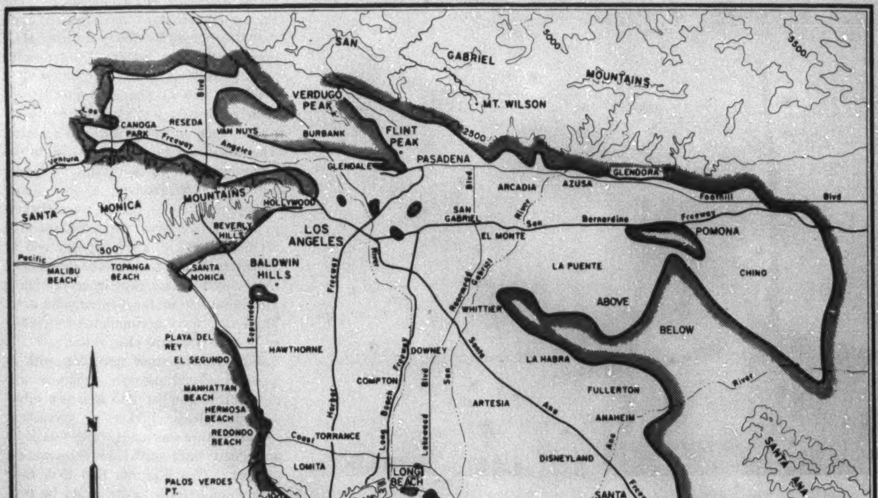
MOBILE-UNIT CONTROL MEASURES

The primary input power to the radio survey equipment was supplied by a Leece-Neville Company heavy-duty alternator. This arrangement provided a heavy-duty primary power source that would not vary when the automobile engine was up to speed required for traffic conditions.

The receiver-recorder was calibrated in the vehicle while the engine was operating at moving-traffic conditions. Before calibration, the equipment was operated for 30 minutes under this environment in order to stabilize circuit elements.

Calibration of the receiver-recorder was accomplished by using a Measurements Corporation standard signal generator, model 560 FM. The mobile equipment was calibrated at the beginning of each day's activities. It is recognized that exacting or laboratory measurements require a continuous check on test-equipment calibration. However, from past experience and in consideration of the

Fig. 8 (below). Plan of 37-dbu-coverage contour from Mt. Wilson


SAN GABRIEL MOUNTAINS
5000
3500
2500
VERDUGO PEAK
MT. WILSON
FLINT PEAK
CANOGA PARK
RESEDA
VAN NUYS
BURBANK
GLENDALE
PASADENA
GLENDORA
ARCADIA
AZUSA
SANTA MONICA MOUNTAINS
HOLLYWOOD
BEVERLY HILLS
LOS ANGELES
SAN GABRIEL
EL MONTE
POMONA
CHINO
LA PUENTE
MALIBU BEACH
TOPANGA BEACH
SANTA MONICA
BALDWIN HILLS
WHITTIER
ABOVE
BELOW
PLAYA DEL REY
EL SEGUNDO
HAWTHORNE
DOWNEY
LA HABRA
MANHATTAN BEACH
HERMOSA BEACH
REDONDO BEACH
COMPTON
ARTESIA
FULLERTON
ANAHEIM
TORRANCE
LOMITA
LONG BEACH
DISNEYLAND
PALOS VERDES PT.
SAN PEDRO
LOS ANGELES LONG BEACH HARBORS
SEAL BEACH
MIDWAY CITY
SANTA ANA
SANTA ANA MOUNTAINS
37DBU CONTOUR
PACIFIC OCEAN
SAN PEDRO CHANNEL
HUNTINGTON BEACH
NEWPORT BEACH
CORONA DEL MAR
N

final results desired it was decided that the test would be conducted with a calibration of equipment at the beginning and at the end of each test run. The receiver-recorder held the required degree of calibration for the necessary data accumulation. After the initial calibration the vehicle engine and radio equipment were operated continuously until final calibration at the end of the day's survey. The calibration of the receiver-recorder at the end of a test run confirmed our past experiences. The slight differences noted in calibration between the beginning and end of each test run was easily adjusted in the data analysis. The signal generator was continuously operated until all further necessity for calibration of survey receivers had passed. This meant that the signal generator operated continuously for approximately 2 weeks.

A simultaneous correlation between the first limiter–last i-f stage current meter, a discriminator voltage meter, equipment supply voltage meter, and the stylus displacement was made during calibration. This correlation provided a service check on the recorder performance during field work with a signal input of 3 μv. Calibration readings were posted at the meters so that frequent checks of equipment performance could be made during data accumulation.

Data observations were made only while the engine rpm were such that the primary supply voltage from the vehicle's electrical system was constant. Recordings on the Varian graphic recorder made under primary low-input-voltage conditions were not summarized in the data analysis.

Base-Transmitter Control Measures

An installation similar to that in the mobile unit was made in a laboratory. The signal from the base station was continuously monitored at this fixed location during the time that field observations were made. This insured that variations of base-transmitter power would be recorded so that compensatory adjustments could be made in the data analysis. Because the radiated energy as measured in the laboratory did not change during the field work, adjustments were not required.

Data Analysis

The voice-recorded observation numbers, time, and location were correlated, and equivalent μv values were assigned to each graphic-recorder observation. This operation involved considerable latitude of judgment because the value assigned was a compromise based first, upon an average of the field-strength variations and second, upon the trend of the variation in that area. During this and previous propagation surveys the short-term variation was observed to exceed the full scale of the recorder. In this case the variations of field strength were in excess of 18 db.

An average was derived from the data and recorded on a suitable map which resulted in the 37-dbu contour. Establishment of the 37-dbu contour required 3,410 separate field observations. Fig. 8 shows a map plan of the 37-dbu contour of the coverage from Mt. Wilson. The coverage line does not surround Mt. Wilson because the survey was limited to the community of interest: Los Angeles and the neighboring cities for which Los Angeles may be regarded as the marketing center. North of Los Angeles the mountain terrain which with Mt. Wilson, is in a national forest, is not crossed by a public-road net. The intent of the service contour was to enclose the service area with a continuous line as far as possible. Therefore the contour line was drawn on the map where the averaged signal level first dropped below the standard 37 dbu on the test runs into the foothills. It was also recognized that further into the mountains there are many areas where the field strength is equal to or greater than the 37-dbu standard. This same procedure was used along the hills separating Santa Monica from Reseda and the hills between Whittier and Pomona. The line between El Toro and Corona Del Mar was placed there because it most suitably joined the limits of coverage on US Highway 101 and 101 Alternate. It is of interest to note that the 37-dbu contour circumscribes the area considered as the limit of reliable telephone service, and also provides protection to our urban MTS 450-mc system from cochannel interference according to part 21 of the FCC's rules.

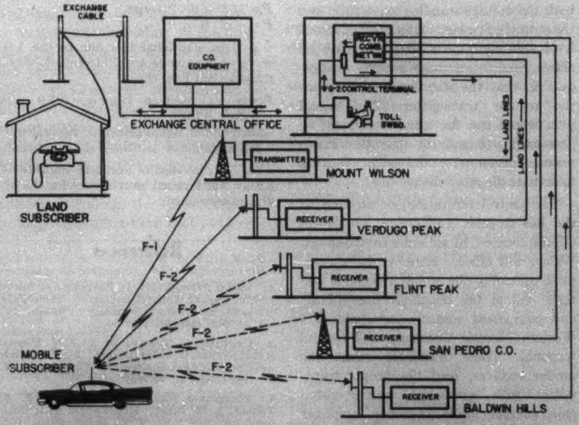


Fig. 9. Los Angeles urban 450-mc common-carrier MTS-equipment arrangement

General Information

Common carrier (MTS) is a semiduplex system employing two different frequencies per radio channel. One frequency is for the base-station transmitter and mobile receiver while the second frequency is for the mobile transmitter and base-station receiver operation. The base station is operated in a full duplex condition, i.e., during the service period the base transmitter comes on the air and remains on the air until the service period is terminated. The mobile transmitter is operated as a push-to-talk (transmitter on) type of operation. One requirement for the continuous operation of the base station and the 2-frequency assignment is a rebroadcast feature which allows a mobile-to-mobile telephone conversation. Another reason for a continuous operation of the base transmitter is to signal any mobile subscriber wishing to use the channel that the channel is busy and already in use. By monitoring the base-station frequency, the complete two-way conversation may be heard when a mobile-to-land telephone station or when a mobile-to-mobile conversation is held. However, when monitoring the mobile transmitting frequency only the mobile transmission is heard. Fig. 9 shows the 450-mc system in Los Angeles.

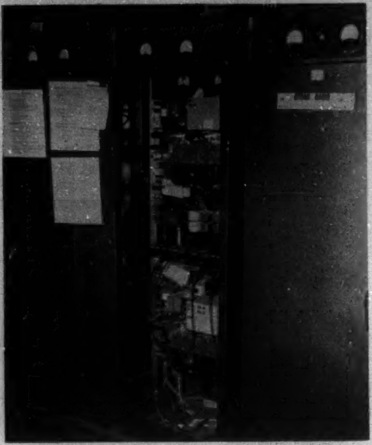


Fig. 10. Regular and standby transmitter with associated patch jack field

At present the 450-mc MTS channels in service are:

Channel	Base Transmitter, Mc	Mobile Transmitter, Mc
QA	454.45	459.45
QB	454.55	459.55
QH	454.75	459.75
QM	454.95	459.95

The base-station equipment is General Electric Company Progress Line model *4ET24A* with an output power of 15 watts or +12 dbw (decibels above one watt). After consideration of transmission-line loss and antenna power gain there is approximately +12 to +13 dbw of effective radiated power. This radiated energy results in a receiver power in the service area of −127 to −89 dbw to the General Electric Progress Line *4ER26B* mobile receiver.

Common-carrier MTS operating tolerances are outlined in part 21 of the FCC's rules as:

1. Frequency stability ±.0005% to the assigned frequency (approximately ±2.3 kc).
2. Measure the operating frequency with equipment capable of ±0.00025% tolerance (approximately ± 1.1 kc).
3. Frequency deviation of ±15 kc (frequency modulation).

For service protection we have installed standby transmitters. This protection is accomplished with the use of one 2-frequency standby transmitter per two regular transmitters. The transmitter change-over is semiautomatic and is, readily placed in service by putting one patch cord up in the 2-frequency transmitter, which connects the telephone line, connects the control function, and actuates the antenna change-over relay to the standby transmitter. To place a standby transmitter in service instead of a regular transmitter takes approximately 2 seconds. In order to replace the standby with regular transmitter the patch is pulled down. Approximate time is 1 second to accomplish this change. Fig. 10 shows the patch jack field.

We plan to make the regular to standby change-over automatic. Two plans are currently being studied. One plan is by the use of M.C. Jones Electronics Company Station Guardian which measures both the voltage standing wave ratio and the output r-f power. At a predetermined level this equipment would actuate an alarm circuit and cause a relay change-over so that the standby would replace the regular transmitter. The second plan would use the same equipment to give an alarm and the operating maintenance personnel would push a button to actuate the relay change-over.

Our control terminals are situated at the Los Angeles Toll Office, 434 South Grand, located 16 air miles from our Mt. Wilson toll office. Here we have complete control of the mobile telephone system. All of the telephone lines (audio and equipment control) terminate in a Western Electric Company *G-2* control terminal. These terminals select the proper receiver, feed the proper transmitter with automatic voice-operated gain control audio-frequency amplifier and afford the maintenance personnel test facilities to properly maintain the system.

Along with system control we maintain a frequency-measuring center where we can measure carrier frequency and frequency deviation to ±0.0001% or approximately ±450 cycles. In our special test center, we are employing unusual techniques which are discussed in a companion paper.[2]

We have all four of the "Q" 450-mc MTS channels loaded up to 60 subscribers per channel. We have applied for four additional 450-mc common-carrier channels which we hope to put into service sometime in 1960. These four additional channels will give some relief to the large demand for this type of telephone service in the Los Angeles metropolitan area.

Appendix

The equation used to determine μv input to the mobile receiver is:

$$P=\frac{\left(\frac{E\lambda}{\pi}\right)^2}{4R}$$

where

E = volts per meter (field strength)
λ = wavelength, meters
R = resistance of 1/2-wave antenna, ohms

If

$E=37$ dbu $=70.79\ \mu v$/meter: $\log E=\bar{5}.84997$
$\lambda=0.6597$ meters: $\log \lambda=\bar{1}.81935$
$R=72$ ohms: $\log 4R=2.45939$
$\pi=3.14$: $\log \pi=0.49715$

in logarithm form the equation becomes:

$$\log P=2\ (\bar{5}.84997+\bar{1}.81935-0.49715)-2.45939$$
$$\log P=\overline{13}.88495$$
$$P=7.673\times10^{-13}\text{ watts}$$
$$P=-121.9\text{ dbw}$$

Adding additional loss due to the 1/4-wave whip antenna and 10 feet of *RG-58/U* transmission line, then:

Power at receiver input = −121.9 dbw −6 db = −127.9 dbw
Closed-circuit signal-generator equivalent = 2.953 μv

Therefore, because of accuracy limitations, 3.0 μv will be used as reference for the 37-dbu measurement.

References

1. RADIO PROPAGATION AT FREQUENCIES ABOVE 30 MEGACYCLES, Kenneth Bullington. *Proceedings*, National Electronics Conference, Chicago, Ill., vol. 35, Oct. 1946, pp. 1122-36.

2. CENTRALIZED TESTING OF A VHF MOBILE RADIO SYSTEM, D. L. MacDonald. *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 79, 1960 (*Paper no. 60-996*).

3. ELECTRONIC MEASUREMENTS (book), F. E. Terman, J. M. Pettit. McGraw-Hill Book Company, Inc., New York, N. Y., second edition, 1943, chap. 11, "Radio Waves," pp. 449–83.

Use of Moon or Satellite Relays for Global Communications

LEANG P. YEH
NONMEMBER AIEE

THE RANGE of global communications has been limited, in the past, by the ionospheric mode of propagation for the very high frequencies, and by the ground-wave mode of propagation for the very low frequencies.

There is a possibility, however, of extending the range by means of moon or satellite relays. This paper investigates this possibility from the practical system performance point of view.

Moon Relay

Range Versus Usable Time

As shown in Fig. 1, the plane of the moon's orbit around the earth is inclined to the plane of the ecliptic (the earth's orbit around the sun) at an angle of about 5.1 degrees, the earth's axis is tipped about 23.5 degrees from the perpendicular to the ecliptic, and the moon's axis is inclined about 6.5 degrees from the perpendicular to the plane of its orbit. In addition, the period of the earth's revolution around the sun is about 365 days, while that of the moon's revolution around the earth is only about 27 days.

Because of these astronomical complexities, the usable time of moon relay, the total time per day during which two stations on the earth's surface can see the moon simultaneously, varies from day to day by amounts depending on the longitudes and latitudes of the two stations. Daily periods during which moon relay is possible at Greenwich meridian can be derived from the times of moonrise and moonset obtained from the "American Ephemeris and Nautical Almanac"[1] for latitudes from +60 degrees to −60 degrees, or from "The Air Almanac"[2] for latitudes including the polar regions. For longitudes other than that of Greenwich, a method of obtaining the local time of moonrise and moonset is given in "Navigation and Nautical Astronomy."[3]

The range of moon relay (great-circle distance) may be computed from the known longitudes and latitudes of the two stations.[4]

A definite relation exists between the range and usable time of moon relay between two earth stations. Simple cases when both stations are located at the equator or at the same longitude will serve as good examples.

The moon-relay geometry at the equator, considering the earth as an idealized sphere, is shown in Fig. 2. If the mean distance between the earth and the moon is taken as 240,000 miles and the earth's radius as 4,000 miles, the angle θ is approximately equal to 89 degrees. Relay between stations A and B has a range of approximately 6,000 miles and a usable time of about 6 hours per day, assuming that both transmitting and receiving antennas are steerable to follow the moon. However, one can trade range for usable time.

The relation between range and usable time is illustrated in Fig. 3. If the two stations (A and C in Fig. 2) are 12,000 miles apart, or approximately half the circumference of the earth, relay is impossible because the two stations can see the moon simultaneously at only one instant. If usable time of, say, 10 hours per day is required, the range is only 2,000 miles.

When the two stations are located at the same longitude, range and usable time are independent. The usable time is about 12 hours for all ranges up to about 12,000 miles, half the circumference of the earth.

In general, therefore, the usable time of moon relay between two earth stations varies from zero to 12 hours per day, and the range varies from zero to 12,000 miles.

Moon Echoes

The first use of radar equipment to obtain echoes from the moon was reported independently by the U.S. Army[5] and Bay[6] of Hungary. Subsequently, experimental work has been carried on in Australia,[7,8] England,[9] and the United States.[11–12] The possibility of obtaining moon echoes is well confirmed today.

The basic radar equation for the median received power for a target subtending an angle greater than the antenna beam width may be expressed as:

$$P_r=\left(\frac{P_t g_t}{4\pi d^2}\right)A_b\left(\frac{A_r}{4\pi d^2}\right) \tag{1}$$

where

P_r = median received power
P_t = transmitted power
d = transmission path length = 240,000 statute miles for moon echo
A_b = effective echo area of the target
A_r = effective area of receiver antenna
g_t = transmitter antenna power gain = $4\pi A_t/\lambda^2$
A_t = effective area of transmitter antenna
λ = wavelength = c/f
c = velocity of light = 3×10^{10} cm (centimeters)/second = 3×10^{8} miles/second
= 186,000 miles/second $\approx 10^9$ feet/second
f = frequency

Expressing both A_r and A_b as power gains,

$$g_r=\text{receiving antenna power gain}=\frac{4\pi A_r}{\lambda^2} \tag{2}$$

and

$$g_b=\text{echo area power gain}=\frac{4\pi A_b}{\lambda^2} \tag{3}$$

Substituting equations 2 and 3 into 1:

$$P_r=P_t g_t g_r g_b\left\{\left(\frac{\lambda}{4\pi d}\right)^2\right\}^2 \tag{4}$$

The factor $(\lambda/4\pi d)^2$, sometimes called the free-space attenuation or loss, is the ratio of received to transmitted power for line-of-sight transmission in free space (a region free of all objects that might absorb or reflect energy) between isotropic antennas (antennas with a power-gain ratio of unity).

For system design, it is more convenient to use decibels (db) instead of power ratios. Expressed in db, equation 4 becomes:

$$10\log_{10}P_r=10\log_{10}P_t+G_t+G_r+G_b-2L_{FS} \tag{5}$$

where

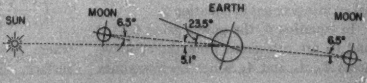


Fig. 1. Relation between earth's and moon's orbit; sizes and distances not to scale

---

Paper 60-1064, recommended by the AIEE Radio Communication Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Pacific General Meeting, San Diego, Calif., August 8–12, 1960. Manuscript submitted April 15, 1960; made available for printing June 15, 1960.

Leang P. Yeh, formerly with the General Electric Company, Philadelphia, Pa., is with Page Communication Engineers, Inc., Washington, D. C.

The work covered by this paper was initiated while the author was with Electronics Division, Westinghouse Electric Corporation, Baltimore, Md., and completed at Aerosciences Laboratory, General Electric Company, Philadelphia, Pa.

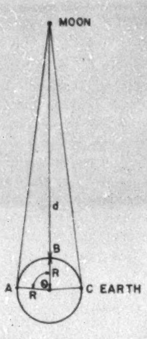


Fig. 2. Moon-relay geometry at equator

$$\cos\theta = \frac{R}{Rd}$$

$R \approx 4{,}000$ miles

$=$ true earth radius

Mean $d \approx 240{,}000$ miles

$\theta \approx 89$ degrees

Relay between A and B:

$$AB = \text{range} \approx \frac{2\pi R}{4}\ 6{,}000 \text{ miles}$$

Time ≈ 6 hours

$10 \log_{10} P_r =$ received power in dbw (db relative to 1 watt)

$10 \log_{10} P_r =$ transmitted power in dbw

$G_t =$ transmitted antenna power gain in db

$G_t =$ transmitted antenna power gain in db

$$= 10 \log_{10} g_t = 10 \log_{10} \frac{4\pi A_t}{\lambda^2} \qquad (6)$$

$G_r =$ received antenna power gain, db

$$= 10 \log_{10} g_r = 10 \log_{10} \frac{4\pi A_r}{\lambda^2} \qquad (7)$$

$G_b =$ echo area power gain in db

$$= 10 \log_{10} g_b = 10 \log_{10} \frac{4\pi A_b}{\lambda^2} \qquad (8)$$

$$L_{FS} = \text{free space loss in db} = 20 \log_{10} \left(\frac{\lambda}{4\pi d}\right)$$

$$= 20 \log_{10} \left(\frac{c}{4\pi d f}\right)$$

$$= -[36.6 + 20 \log_{10} d_{mt} + 20 \log_{10} f_{mc}] \qquad (9)$$

where

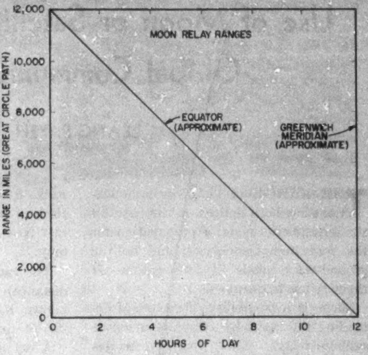
$d_{mt} =$ transmission path length, statute miles

$f_{mc} =$ frequency, megacycles (mc) per second

It should be noted that free space loss bears a negative sign as shown in equation 9. This negative sign has been taken into account in equation 5.

In equation 5, the power gains G_t, G_r, and G_b are calculated from the effective areas A_t, A_r, and A_b, which are sometimes quite different from the physical areas. For parabolic antennas with circular cross section and taper illumination (intensity at edge 10 db down from that at center), the effective area is approximately 0.64 of the actual area, and the gain becomes:

Fig. 3. Range versus usable time



$$G = 10 \log_{10} \frac{4\pi}{\lambda^2} \left(\frac{0.64\pi D^2}{4}\right)$$

$$\approx 10 \log_{10} 2\pi \left(\frac{D}{\lambda}\right)^2 \qquad (10)$$

where $D =$ diameter of antenna aperture. The half-power (3-db) beam width or beam angle of a parabolic antenna with taper illumination is given by:

$$\alpha = \text{beam width} \approx 70 \frac{\lambda}{D} \text{ degrees} = 1.22 \frac{\lambda}{D} \text{ radians} \qquad (11)$$

For the target, the moon in this instance, it has been shown by Norton et al[13] that the effective echo area would be equal to the projected area if the moon were an idealized body whose surface were smooth, perfectly reflecting, and spherical.

Effective echo area of ideal moon

$$= \pi\left(\frac{\text{Moon's diameter}}{2}\right)^2 = \pi(1080)^2$$

$$= 3.66 \times 10^6 \text{ square miles} \qquad (12)$$

Since the moon is far from an idealized body, the effective echo area must be modified accordingly.

First, the moon's surface is not perfectly reflecting. The power-reflection coefficient, p, of any surface is a function of its dielectric constant and may be expressed as:

$$p = \left(\frac{1-\sqrt{E}}{1+\sqrt{E}}\right)^2,\ p \leqq 1 \qquad (13)$$

where $E =$ dielectric constant of moon's surface. Several values of power reflection coefficient have been postulated a shown in Table I.

Second, the moon's surface is not perfectly smooth. The roughness of the surface scatters radiation in all directions, tending to increase the echo cross section by a so-called directivity factor, q. Again different values have been postulated for this factor. Grieg[14] gives a q value of 8/3. Here the assumption is made that Lambert's cosine law of reflection may be applied to a rough or diffuse moon. Kerr's[7,8] value is 5 for 20 mc. His reasoning is that, since optical directivity of the moon's surface to sunlight is 5.7, it should be smaller for radio wavelengths. A value of 5 is therefore suggested.

Because the energy is returned from the whole of the visible hemisphere of the rough moon and not only from the first Fresnel Zone[5] as in the case of a smooth surface, the improvement in directivity is complicated or counteracted by two other effects:

1. Pulse elongation effect. The duration of the returned pulse is increased by the time T, required to travel forward and back along the radius of the body. In the case of the moon,

$$T = \frac{2 \times \text{moon's radius}}{\text{velocity of light}} = \frac{2 \times 1{,}080}{186{,}000} = 0.0116 \text{ second}$$

Table I. Moon's Power-Reflection Coefficient, p

	Mofenson,[5] Dewitt[10]	Kerr[7,8]	Grieg[14]
Type of surface		dry rock	rock
Dielectric constant	6	5	5
Reflection coefficient	0.18	0.15	0.15 (0.10*)

* A value of 0.10 is recommended by Grieg.

Table II. Frequency Spread, f_p, Maximum

Wavelength, Meters	Frequency, Mc	Frequency Spread, Cps
10	30	2
1	300	20
0.1	3,000	200

Grieg[14] has found that power reduction factor, m, due to pulse elongation effect may be expressed as:

$$m \approx \frac{t}{T} \text{ for } t<T \text{ only; } m<1 \tag{14}$$

where t=pulse length. If $t \geqq T$, there is no power reduction.

The pulse elongation time, T, of 0.0116 second has been challenged recently by Senior et al,[15] who used a "quasi-smooth" moon theory. On the basis of experimental results, they estimated that the active echo area is within a region less than one third of the moon's radius in extent, and that the maximum pulse elongation is thereby reduced from 11.6 ms (milliseconds), corresponding to the full depth of the moon, to no more then 600 μsec (microseconds). As will be seen later, this concept seems to agree with some of the moon-echo experiments[16,17] using short pulses in the United States.

2. Frequency spread effect. If the body is rotating, there is a frequency spread, f_p, in the spectrum of the echo proportional to the velocity, v_s, of the moon. According to Pawsy et al,[18] this maximum frequency spread, f_p, is equal to:

$$f_p = \frac{4v_s}{\lambda} \tag{15}$$

where

$$v_s = \frac{\text{circumference of moon's body}}{\text{rotation period}}$$

$$= \frac{\pi(2\times 1{,}080)\text{miles}}{27^1/_3 \text{ days}}$$

$=2.8\times 10^{-3}$ miles per second $=4.8$ meters per second

λ = wavelength

Table II shows the magnitudes of f_p at various wavelengths.

If the bandwidth, B, of the receiver is less than this frequency spread, only a part of the energy is received. The power reduction factor, s, due to frequency spread effect may be expressed as[18]

$$s = \frac{4B}{\pi f_p} \text{ for } B<f_p \text{ only; } s<1 \tag{16}$$

If $B \geqq f_p$, there is no power reduction.

In most experimental setups, the receiver bandwidth, B, would be larger than the frequency spread, f_p, as indicated in Table II. The frequency spread effect will therefore be neglected.

Summing up, the effective echo area, A_e, becomes:

$$A_e = \pi\left(\frac{\text{Moon's diameter}}{2}\right)^2 pqm \tag{17}$$

where

$$p = \text{power reflection factor} = \left(\frac{1-\sqrt{E}}{1+\sqrt{E}}\right)^2 \qquad p \leqq 1$$

q = directivity factor $\qquad q \leqq 1$

m = pulse elongation factor $= t/T \qquad m \leqq 1$

A summary of the system parameters and measured and predicted received signal powers of the major moon-echo experiments are shown in Table III. In this table, the general parameters and measured received signals are taken from the original papers; the effective echo area of the moon is calculated from equation 17, using $p\times q=0.1$ and $T=10^3$ μsec; the predicted received signal is calculated from equation 5; the receiver sensitivity or the receiver thermal noise level, N, is calculated from the following:

$$N = \text{receiver sensitivity} = FKTB \text{ watts} \tag{18}$$

where

F = effective receiver noise figure

K = Boltzmann's constant $=1.38\times 10^{-23}$ joules/K (degrees Kelvin)

T = 290 K, ambient temperature at the earth's surface

B = receiver bandwidth

$KT = 1.38\times 10^{-23}\times 290$
$= 4\times 10^{-21}$ watts $= -204$ dbw

The measured circuit loss, L_c, is the total transmission loss of the two-way radar or passive-relay system between isotropic transmitter and receiver antennas:

$$L_c = 10\log_{10} P_r - 10\log_{10} P_t - G_t - G_r \tag{19}$$

where P_r = measured received signal power.

Table III. Moon-Echo Experiments

System	Mofenson,[9] Dewitt[10]	Kerr, et al[7,8]	Sulzer, et al[11]	Browne, et al[9]	Trexler[16]		Yaplee[17]	Fricker[12]
A. General Parameters:								
Type of signal	pulse	pulse	cw	pulse	pulse	cw	pulse	cw
Frequency, mc	111	21	418	120	198	198	2,860	400
Peak power, kw	3	70		10	1,200		2,300	
Pulse width, seconds	1/4	1/4		0. 3	$10\times 10x^{6}$		$2\times 10x^{6}$	
Pulse repetition frequency, pulses per second	1/4	1/6		0.60	10		250	
Average power, kw			20			0.1		40
Transmitting antenna gain, db	24	20	21	25	40	40	51	21
Receiving antenna gain, db	24	17	30	25	40	40	51	36
Transmitting antenna beam width, degrees	15			12	1.4	1.4	0.5	
Receiving antenna beam width, degrees	15			12	1.4	1.4	0.5	
B. Receiver Sensitivity:								
Receiver bandwidth, cps	100	70	10^3	26	10^5	3×10^3	35×10^4	50
Receiver bandwidth, db	20	18	30	14	50	35	55	17
Receiver noise figure, db	5	20	4	5	4	4	8	6
$K\,T$, dbw	−204	−204	−204	−204	−204	−204	−204	−204
Receiver sensitivity, dbw	−179	−166	−170	−185	−150	−165	−141	−181
C. Effective Echo Area:								
Projected area, square miles	3.66×10^6	3.66×10^6	3.66×10^6	3.66×10^6	3.66×10^6	3.66×10^6	3.66×10^6	3.66×10^6
Reflection times directivity, pq	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10
Elongation, m					10^{-2}		2×10^{-2}	
Effective area, square miles	3.66×10^5	3.66×10^5	3.66×10^5	3.66×10^5	3.66×10^3	3.66×10^5	7.32×10^3	3.66×10^5
D. Predicted Median Receiver Signal:								
Transmitter power, dbw	+35	+48	+43	+40	+61	+20	+64	+46
Transmitting antenna gain, db	+24	+20	+21	+25	+40	+40	+51	+21
Receiving antenna gain, db	+24	+17	+30	+25	+40	+40	+51	+36
Echo-area gain, db	+122	+108	+134	+123	+108	+128	+124	+134
Free-space loss, db	−370	−341	−393	−371	−380	−380	−426	−392
Received signal, dbw	−165	−148	−165	−158	−131	−152	−136	−155
E. Measured Received Signal:								
Maximum received signal, dbw	−159	−140	−160					
Average received signal, dbw	−168	−150	−165	−157	−130	−152	−138	−155
Minimum received signal, dbw	−177	−160	−170					
F. Measured Average Circuit Loss:								
Circuit loss, db	−251	−235	−259	−247	−271	−252	−304	−258

It can be seen that the predicted median received signal levels agree surprisingly well (within ±2 db) with measured values in all cases. The only losses not taken into account in the predicted calculations are those due to the feed and coupling in the antenna circuits. However some of these losses may have already been included in the consideration of the antenna gains. In any event, such losses probably do not exceed 2 to 3 db.

Some conclusions may be drawn from Table III:

1. For continuous wave (cw) and long pulse signals (1,000 μsec or longer), the effective echo area of the moon may be considered as 0.1 of the projected area. For shorter pulses, the pulse elongation effect should also be considered.
2. A circuit loss of 250 db between isotropic antennas for 100 mc, and a 6-db increase of loss per octave increase of frequency may be used for future moon-echo system design using long pulses or cw. For pulse lengths shorter than 1,000 microseconds, an additional power reduction factor, t/T, should be used; t=pulse length in μsec and T=1,000 μsec. Fig. 4 shows the moon-circuit loss for various frequencies, and Fig. 5 the power-reduction factor in db for various short-pulse lengths.

SYSTEM CONSIDERATIONS

Two types of moon relay are possible: using the moon as a passive reflector, and using the moon as an active repeater station. At present, only passive relay is feasible; active relay must wait until equipment can be set up on the moon.

The use of the moon as a passive reflector for earth-to-earth relay has been thoroughly analyzed theoretically by Grieg et al.[14] It was concluded that the most important consideration for such a system is the moon's surface. If the moon is smooth, there is no limitation to the type of transmission possible, telegraphy, voice, or wideband. On the other hand, if the moon is rough, essentially only narrow-band telephony and telegraphy can be accommodated.

Experiments over many decades have concluded that the moon's surface is rough for optical frequencies. Whether this is also true for radio frequencies is highly debatable and cannot be settled until more extensive experimental data are available. It is most probable that the moon's surface is neither smooth nor rough, but a mixture of both.

In any communication system the minimum usable carrier-to-total noise-power ratio (the threshold), is of utmost importance. This ratio determines whether the input signal can be detected in the presence of noise associated with the receiver input, including the antenna system. This is the first limitation on the system performance.

In the region of 100 to 10,000 mc, considered feasible for moon relays, the thermal noise inherent in the receiver predominates at isolated locations where the receiver terminals are likely to be located. The following discussion is limited to thermal noise only.

The minimum usable carrier-to-noise power ratio depends on the modulation techniques used. For practical system design, a value of 15 db seems to be an average acceptable figure.

The received carrier power, C, in dbw may be expressed as:

$$C=10\log_{10}P_t+G_t+G_r-L_c-L_F-L_L \qquad (20)$$

where

C=received carrier power, dbw
$10\log_{10}P_t$=transmitter power, dbw
G_t=transmitter antenna power gain, db
G_r=receiver antenna power gain, db
L_c=moon-circuit loss, db, from Figs. 4 and 5
L_F=fading allowance, db
L_L=other losses in db such as feed and coupling losses in antenna systems, etc.

The thermal noise power, N, in dbw may be obtained from equation 18:

$$N=10\log_{10}(FKTB) \qquad (21)$$

The carrier-to-thermal-power ratio in db then becomes:

$$\frac{C}{N}=10\log_{10}P_t+G_t+G_r-L_c-L_F-L_L-10\log_{10}(FKTB) \qquad (22)$$

From equation 22, system parameters could be determined.

FADING AND SYSTEM RELIABILITY

Of the parameters in equation 22, only the fading allowance, L_F, is yet to be established. The moon-echo experiments indicate that the received signals are subject to continuous fading of 20 db or more at both slow and fast rates. The slow fading, with a period of an hour or less, is believed due to the Faraday effect (the plane of polarization of the radio waves is rotated as they traverse the ionsphere in the presence of the earth's magnetic field). This effect has been reported to vary inversely between the first and second power of frequency. Above 400 mc the effect seems negligible.

The fast fading, with a period of seconds or longer, is believed due to the moon's libration (an apparent oscillation of the moon's face, as seen from a point of the earth, arising from the diurnal rotation of the observer on the earth and from variations in the moon's orbit). Browne et al[9] have analyzed 2-minute sample recordings and found the fading distribution to be very close to Rayleigh's equation. Moreover, atmospheric refraction in the troposphere and multipath effects may also have some bearing on the fading of the received signal. Therefore the fading characteristic is a combination of many effects.

In general, a fading margin of 10 to 20

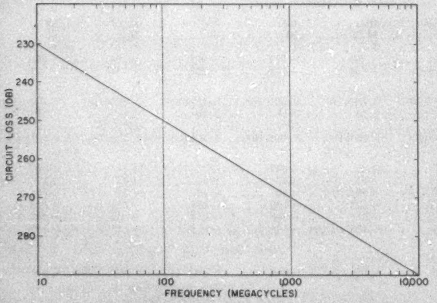


Fig. 4. Moon circuit loss

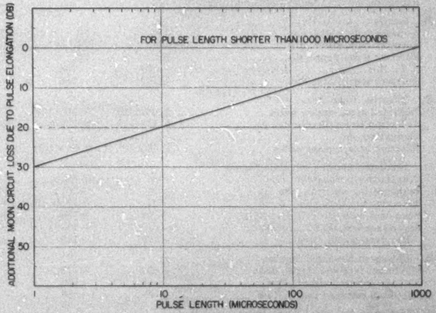


Fig. 5. Pulse elongation effect

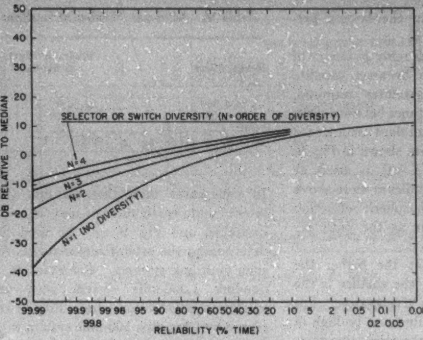


Fig. 6. Fading allowance and reliability

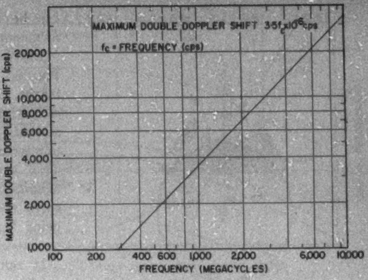


Fig. 7. Maximum moon-relay Doppler shift

db or even more should be allowed for a workable system. Extensive signal-level measurements on a long-term basis are required to determine the more reliable statistical fading characteristics. In the present study, a Rayleigh fading distribution as shown in Fig. 6 is assumed. From this distribution curve, the reliability of the system can be determined if the fading allowance is known. At the present state of the art it is not expected that a moon-relay system would be able to have a reliability of more than 99%. A fading allowance of 20 db, therefore, is recommended; see the Rayleigh curve in Fig. 6. At 99% reliability, the fading is approximately 20 db.

Fading may be reduced by diversity techniques as indicated in Fig. 6, where fading distributions for several orders of diversity are also plotted. Fig. 6 is reproduced from another paper by the author.[19]

DOPPLER SHIFT

Another factor worth considering is the Doppler shift of the carrier frequency resulting from the relative motion between the moon and the earth. For a round trip betwen earth and moon there is a double Doppler shift. The magnitude of the Doppler shift depends on the location of transmitting and receiving terminals, the frequency of operation, and the month and hourly period. A method of calculating this double Doppler shift is given by Grieg et al.[14] The maximum that can be expected at various frequencies is given by the following:

d_s = maximum double Doppler shift, cps (cycles per second)

$$\approx 3.5f \times 10^{-6} \text{ cps} \qquad (23)$$

where f = carrier frequency, cps.

Equation 23 is plotted in Fig. 7. It can be seen that for an upper uhf (ultrahigh-frequency) system, say 10,000 mc, the double Doppler shift may be as high as 35 kc. Although smaller Doppler shifts have been successfully compensated by several practical schemes, such large shifts may be a serious problem.

SAMPLE CALCULATIONS

The choice of frequency to be used for space communications has been discussed in another paper.[20] It was found that the frequency ranges are as shown in Table IV.

A sample calculation for a typical 2,000-mc system is shown in Appendix I. The system will, therefore, work 80% of the time (see Fig. 6). For a reliability of 99% some means, such as using a diversity system or increasing the power and antenna size, must be used.

Satellite Relays

ORBITAL REQUIREMEMTS[21–24]

Since the satellite must travel in the centrally directed force field of the earth, its orbit lies in a plane passing through the center of the earth. There are three possible types of orbital planes:

1. Circumpolar, passsing through the two poles of the earth.
2. Equatorial, passing through the earth's equator.
3. Inclined, at an angle to the earth's equator. If the angle of inclination is equal to the ecliptic angle, it is sometimes called the ecliptic plane.

Inasmuch as the orbital period is much shorter than that of the earth's rotation, the orbital path described on the earth will not retrace itself in successive revolutions, but will look like a spiral inscribed on a sphere, so that even a circumpolar path will not lie exactly on a meridian. However, there is one exception, the case of equatorial orbit. When a satellite is in an equatorial orbit, it will remain at all times above the equator. This is the most ideal condition for satellite relays.

The shape of the orbital path is also very important. The path may be a "great circle," with the center of the earth as its center; or, more generally, a "great ellipse" with the earth's center as one of its focuses. Whether the path is circular or elliptical depends on the injection conditions, primarily the injection velocity and angle. There is only one combination of injection velocity and angle that will cause the satellite to orbit in a circular path. Any slight errors will put the satellite in an elliptical orbit.

Another interesting aspect is the lifetime of the satellite. The most significant parameters that affect the lifetime are the frontal area density of the satellite, expressed as mass (in pounds) per-unit frontal area (in square feet); and the atmospheric density, expressed as slugs per cubic feet (slug = pound/32.2). It is obvious that a large satellite at too low an altitude will not stay up long. Methods[24] have been proposed to calculate the lifetime of satellites in circular or elliptical orbits. Due to the difficulty in estimating accurately the atmospheric densities at altitudes of hundreds or thousands of miles, the calculated life is bound to differ from practical results.

Table IV. Optimum Frequency

Antenna System	Frequency Range, Mc
Isotropic–Isotropic	40– 100
Isotropic–Directive	100– 300
Directive–Directive	2,000–4,000

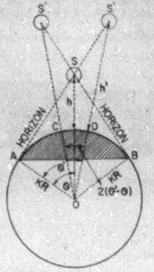


Fig. 8. Satellite-relay geometry, circular equatorial orbit. Shaded area is spherical cap on earth's surface within which line-of-sight communication from satellite will be received

h = height when horizon at A and B meet

$$= \frac{KR}{\cos\theta} - KR$$

R = true earth radius = 4,000 miles

K = effective earth radius factor

$$\theta = \cos^{-1}\frac{KR}{KR+h} \text{ radian}$$

$$h' = \text{satellite height} = \frac{KR}{\cos\theta'} - KR$$

$$\theta = \cos^{-1}\frac{KR}{KR+h'} \text{ radian}$$

$$n = \frac{2\pi}{2(\theta'-\theta)} = \frac{\pi}{\theta'-\theta}$$

n = number of satellites required

$AB = 2KR \times \theta$ = range

$S'A = (KR+h')\sin\theta'$

= maximum slant distances

RELAY GEOMETRY

As discussed in the previous section, the most ideal conditions for relay occur when the satellite is in a circular equatorial orbit. Practically, it may be difficult to fulfill this condition, but the closer this ideal is approached, the better is the performance of the system predicted.

Fig. 8 shows the ideal relay geometry. When the satellite is at a height, h, where the horizons at earth stations A and B meet, relay is impossible because the two stations can see the satellite simultaneously at one instant only. If the satellite is in orbit at a height, h', instead of h, it is seen simultaneously by the two stations through an angle $2(\theta'-\theta)$. Therefore, if there are $n=(2\pi)/(2(\theta'-\theta))=(\pi)/(\theta'-\theta)$ satellites, each at an angular distance of $\theta'-\theta$ from the other, an uninterrupted relay is possible provided that the satellites are equidistant from each other in the same circular equatorial orbit. Such conditions cannot be met in practice, but slight orbital deviations can be tolerated from the system performance point of view.

Based on the satellite relay geometry of Fig. 8, the relations between satellite height, number of satellites required, ranges (great-circle distance between earth stations), and maximum slant distance for uninterrupted relays are shown in Fig. 9. True earth radius ($K=1$) is used at heights of hundreds of miles or more above earth, the effect of atmospheric refraction being negligible as far as the path geometry is concerned.

It can be seen that the higher the satellites are in orbit, the smaller is the number required for a given range. For long ranges, the satellite must be high in orbit. Furthermore, the minimum number of satellites required depends on the range. Table V gives the numbers required for various ranges.

As discussed before, the minimum satellite height for a given lifetime depends on the atmospheric density and satellite size and mass. A height of 300 miles is considered sufficient to maintain a fair sized satellite in orbit for quite some time.

One interesting parameter is the maximum slant distance, the distance which the relay system is expected to cover. It may be several times larger or smaller than the range (the great-circle distance between the two earth stations), depending on the number of satellites used. For instance, as shown in Fig. 9, a 1,000-mile range requires a maximum slant distance of 2,000 miles with ten satellites and 5,000 miles with four satellites, but only about 500 miles with 300 satellites. The smaller the number of satellites, the longer is the maximum slant distance compared with range. This puts rather stringent requirements on some of the vital parameters of the relay system, such as transmitted power and antenna sizes or gain, because of the longer distances to be covered.

A satellite-relay system between two earth stations (earth-satellite-earth) is a simple two-hop link in radio terminology.

Table V. Minimum Number of Satellites

Range, Miles	Minimum Number of Satellites
Up to 4,000	3
5,000–6,000	4
8,000	6
10,000	11

In some cases, more than one link with intermediate earth stations may be more desirable, and Fig. 9 may be used for determining the orbital requirements for such multilink systems. For example, a one-link 1,000-mile system with ten satellites, each 430 miles high, may be replaced by a two-link 500-mile system with an intermediate earth station, with either ten satellites 300 miles high or eight satellites 430 miles high.

For uninterrupted relay, the satellite must be tracked continuously by the two earth stations. In practice, the relay is intermittently interrupted because of the time required to return the antenna to its initial position for tracking the next succeeding satellite, after the first has been tracked to its disappearance below horizon. For a large antenna, the interruption may be quite long because of the large mass involved.

Although an uninterrupted satellite relay system generally requires a large number of satellites, a few schemes may be used to improve the situation:

1. Four equidistant satellites may be placed in a circular equatorial orbit at a height of 22,500 miles (approximately 26,500 miles from the center of the earth). At this height, the satellites appear to be stationary relative to the earth and can serve as relay stations between any two locations on earth, using one-link or multilink arrangements, depending on the range. As discussed previously however, use of a small number of satellites considerably extends the slant distance to be covered, increasing the requirements of some of the vital system parameters.

2. The number of satellites might be reduced without changing the spacing between them, reducing the usable time accordingly. Consider the 1,000-mile-range

Table VI. Satellite Velocity, Orbital Period, and Doppler Shift

Satellite Height, h', Miles	Satellite Velocity, V: Miles/Hour	Satellite Velocity, V: Feet/Second*	Orbital Period, T, Hours	Single Doppler Shift, f_m, Cps: f = 100 Mc	Single Doppler Shift, f_m, Cps: f = 1,000 Mc	Single Doppler Shift, f_m, Cps: f = 10,000 Mc
10	17,800	27,200	1.4	2,620	26,200	262,000
100	17,610	25,800	1.5	2,590	25,900	259,000
1,000	16,000	23,500	1.9	2,350	23,500	235,000
10,000	9,510	14,000	9.1	1,410	14,100	141,000
22,500†	7,600	11,200	24.0	1,020	10,200	102,000
100,000	3,520	5,170	184.2	519	5,100	51,000
1,000,000	1,130	1,660	5,500.0	170	1,700	17,000

* 1 mile/hour = 1.47 feet/second.

† At this height, the orbital period is 24 hours. The satellite appears stationary relative to earth.

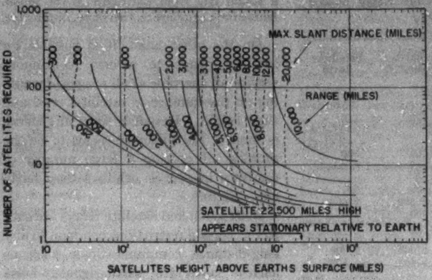


Fig. 9. Number, height, range, and slant distance of satellite

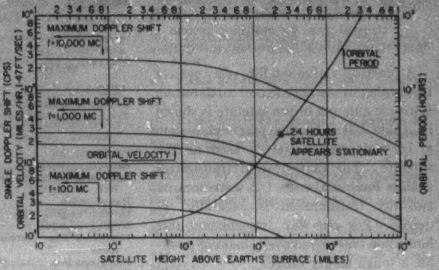


Fig. 10. Doppler effect, orbital velocity, and period

10-satellite system as an example. If five satellites were used instead of ten and the same spacing maintained between satellites (36 degrees of arc in this case), the relay would work half of the time.

3. Fixed (nontracking) antennas might be used. As the satellite passes over the first earth station, it receives and stores the message. As it passes over the next earth station, a common or trigger signal causes the satellite to transmit the stored message. The usable time depends on the antenna beam width of the earth station and is, in general, quite limited since widening the beam for more usable time decreases the gain of the antenna. In order to relay a large amount of information, therefore, a fast rate of transmission must be used. This scheme has been demonstrated recently by the United States "score satellite," a 8,750-pound intercontinental ballistic missile (Atlas) with a 4-minute message-storage time.

SYSTEM DESIGN

Although available data are still sketchy, the radio observations of satellites indicate that the satellite signals like moon echoes, are subject to continuous fading at both slow and fast rates. As indicated earlier, the slow fading is due to the Faraday effect. The fast fading probably results from the spinning of the satellite. Through use of higher uhf and the recently advocated satellite stabilization technique, the fading may be reduced considerably. This, however, must be confirmed by extensive experiments. In the present study, the same magnitude of moon relay fading allowance is recommended for satellite relay.

Since the satellite is a moving object, the carrier frequency is subject to a double Doppler shift for a two-hop link. The maximum single Doppler shift in cps, f_{ss}, may be calculated by:

$$f_{ss}=\frac{fV}{c} \tag{24}$$

where

f = carrier frequency, cps

V = velocity of satellite relative to earth, miles per hour

c = velocity of light, miles per hour = 186,000 ×3,600 miles per hour

The velocity of the satellite is proportional to its height above earth's surface. For a circular path, the velocity may be expressed[21] as:

$$V=R\sqrt{\frac{g}{R+h'}} \text{ miles per hour}$$
$$=\frac{1.13\times10^{6}}{\sqrt{(4000+h')^{1/2}}} \text{ miles per hour} \tag{25}$$

where

R = earth's true radius, miles = 4,000 miles

h' = satellite height above earth's surface, miles

g = 32.3 feet/second^2 = 79,000 miles/hour^2

The orbital period, T, the time required for the satellite to complete one orbit of the earth, may be expressed as:

$$T=\frac{2\pi(R+h')}{V} \text{ hours} \tag{26}$$

Combining equations 24 and 25:

$$f_{ss}=\frac{f}{600(4{,}000+h')^{1/2}} \text{ cps} \tag{27}$$

Combining equations 25 and 26:

T = orbital period, hours

$$=5.5\times10^{-6}(4{,}000+h')^{3/2} \tag{28}$$

Satellite velocities, V, single Doppler shifts, f_{ss}, and orbital period, T, for various satellite heights are tabulated in Table VI and plotted in Fig. 10. Both passive and active relays are feasible. For an active relay, the carrier-to-thermal-power ratio may be expressed as:

1. Earth-to-satellite:

$$\frac{C_s}{N_s}=10\log_{10}P_t+G_t+G_r-L_{FS}-L_F-L_L-10\log_{10}(FKT_sB) \tag{29}$$

where

C_s = received carrier power, dbw, at the satellite

N_s = thermal noise power, dbw, at the satellite = 10 $\log_{10}$ $(FKTB)$

P_t = transmitted power, dbw, at the earth station

G_t = transmitted antenna power gain, db earth station

G_r = received antenna power gain, db satellite

L_{FS} = free-space loss, db (calculated on the basis of maximum slant distance rather than the range)

L_F = fading allowance, db

L_L = other losses in db such as feeder and coupling losses in antenna systems, etc.

F = effective receiver noise figure

k = Boltzmann's constant = 1.38×10^{-23} joules/K

T_s* = ambient temperature of the satellite (assumed the same as the earth station) ≈ 290 K

B = receiver bandwidth

$KT_s=1.38\times10^{-23}\times290—4\times10^{-21}$ watts = −204 dbw

2. Satellite-to-earth:

$$\frac{C_E}{N_E}=C_s+G_s+G_t'+G_r'-L_{FS}-L_F-L_L-10\log_{10}(FKT_EB) \tag{30}$$

where

C_E = received carrier power, dbw, at earth station

N_E = thermal noise power, dbw, at the earth station = 10 $\log_{10}$ FKT_EB

G_s = gain at the satellite repeater

G_t' = transmitted antenna power gain, db (satellite)

G_r' = received antenna power gain, db (earth station)

T_E = ambient temperature at earth station = 290 K

L_F = fading allowance, db

---

* The ambient temperature of the satellite is a rather controversial issue. Methods[35, 36] have been proposed to calculate the skin temperature under various assumptions, but they lead to widely different conclusions concerning the mean temperature. However, it has been reported[37] that the skin temperature of the American satellite Explorer I was found to vary from 270 K to 340 K, and the inside temperature from 100 F (degrees Fahrenheit) to 120 F [40 C (degrees centigrade) to 50 C, 313 K to 323 K]. In the present case, 290 degrees is assumed.

$10\log_{10} M$ = degradation in performance in db due to power summation of thermal noise for multihop system

M = number of hops = 2 (in this instance)

A practical active satellite relay system using a high-gain antenna on earth but an isotropic antenna in the satellite is shown in Table VII. In this practical system a 1/2-watt transmitter is used in the satellite. The earth-station transmitter is thus limited also to 1/2 watt because two-way transmission is generally required. For one-way transmission this limitation will be removed, making possible a much longer range.

It is interesting to note that four times (+6 db) more transmitter power and four times (+6 db) satellite-repeater power gain is required if the antenna size is reduced to one half. Conversely, if the antenna size is doubled, both transmitter power and satellite-repeater power gain can be reduced to one quarter (−6 db).

A typical passive system with 15 satellites 460 miles high and using the same frequency, transmitter power, and antenna size as the typical moon system discussed earlier (see Table IV), is shown in Appendix II.

Assuming that the satellite is a sphere with an effective echo area equal to 0.5 of the projected area, then:

$$10\log_{10}\frac{4\pi A_b\times 0.5}{\lambda^2}=62=\text{echo-area gain of satellite}\qquad(31)$$

where

A_b = projected area of satellite $=\pi r^2$

r = satellite radius

λ = wavelength = 0.5 foot (at 2,000 mc)

From equation 31:

$$\frac{4\pi A_b\times 0.5}{(0.5)^2}=1.58\times10^6$$

$$\frac{4\pi(\pi r^2)\times 0.5}{(0.5)^2}=1.58\times10^6$$

$$r\approx 140\text{ feet}$$

A sphere 280 feet in diameter is required to give the same performance as the moon relay for a relay range of 2,000 miles. Unless inflated devices are used, it is doubtful whether it is possible to put a sphere 280 feet in diameter in orbit at this time.

Moreover, the angles subtended by a 280-foot satellite at a distance of 2,000 miles is very small, on the order of 10^{-3} degrees, and unless its orbit can be predicted with fair accuracy, radio tracking will not be easy. Fortunately, from the experience gained in the observation of the Russian and United States satellites, determination of satellites' orbits is no longer considered a serious problem.

Table VII. Active Satellite Relay System, Two-Way

System	Earth–Satellite	Satellite–Earth
Satellite repeater gain, db		+121
Transmitter power, watts	1/2	1/2 (+121−124 = −3 dbw)
Frequency, mc	300	300
Satellite antenna	isotropic	isotropic
Earth antenna (parabolic), feet	60	60
Range, miles	2,000	2,000
Satellite height, miles	460	460
Number of satellites	15	15
Maximum slant distance, miles	2,000	2,000
Free space loss (L_{FS}), db	−153	−153
Transmitter antenna gain (G_t), db	+ 33	0
Receiver antenna gain (G_r), db	− 0	+ 33
Line loss (L_L), db	− 1	− 1
Net transmission loss, db	−121	−121
Transmitted power (P_t), dbw	− 3	− 3
Median received power (C), dbw	−124	−124
Receiver noise figure (F), db	+ 6	+ 6
Thermal noise per cycle (KT), dbw	−204	−204
Receiver bandwidth (B= 3 kc), db	+ 35	+ 35
Thermal noise power (N), dbw	−163	−163
Multihop degradation, db		+ 3
Net noise power, dbw	−163	−160
Median C/N, db	39	36
Minimum usable C/N, db	15	15
Fading allowance, db	24	21
Reliability, %	99+	99

Again referring to Fig. 9, it can be seen that there is little possibility that the size of the satellite can be substantially reduced by increasing the number of satellites. For instance, if 200 satellites are used for a range of 2,000 miles, the maximum slant distance is reduced to 1,000 miles. Since the size of the satellite required is proportional to the square of the slant distance, it is reduced to $(1{,}000/2{,}000)^2\times280$ or 70 feet, which is still quite large. Therefore, an active satellite relay has a better chance of competing with moon relay at the present time.

Conclusions

Both moon and satellite relays are feasible means for the extension of global communications. In the present state of the art, moon relay seems to have some edge over satellite relay, and since the moon is free for all, moon relay may be the cheapest way to extend the range of global communication. There are, however, certain limitations that must not be overlooked.

1. While the usable time of moon relay is continuous, it is limited and varies from day to day, and depends on the locations of the earth stations and the great-circle distance between them. In practice, a compromise between usable time and range must be observed.
2. It takes 2.5 seconds for a signal to take a round trip from the earth to the moon. Immediate response is, therefore, impossible.

On the other hand, it is possible for satellite relays to achieve continuous service all the time at all ranges, depending on the number of satellites used and their height. Response time, in general, is short. The difficulty seems to be the problem of putting a large number of satellites in precise orbits with a fairly long lifetime.

Both moon and satellite relays are subject to certain restrictions common to any long-distance communications, such as narrow bandwidth, small signal-to-noise ratio, and low system reliability. Therefore, intensive research and development are required on high-power r-f (radio-frequency) transmitters, high-gain antennas for both satellites and earth stations, low-noise amplifiers, bandwidth reduction, novel modulation techniques, etc.

Satellites have the advantage of being readily controllable while the moon is not. Although it may take some time before this advantage can be fully realized, the research and development of satellite relays should not be overlooked in the future planning of global communication systems despite their present limitations.

Appendix I. Typical Moon-Relay System

Parameter	Value
Frequency	2,000 mc
Transmitter power	10kw cw
Antenna size	60-foot parabolic
Antenna beam width	0.6 degree
Transmitter power, P_t	+ 40 dbw
Antenna gain, G_t+G_t	+ 100 db
Median circuit loss, L_c	− 276 db
Line losses, L_l	− 4 db
Median received power, C	− 140 dbw
Noise figure, F	9 db
Noise power per cycle (kT)	− 204 dbw
3-kc bandwidth, B	35 db
Noise power, N	− 160 dbw
Median C/N	20 db
Minimum usable C/N	15 db
Fading allowance, L_F	5 db
Reliability	80% (no diversity) 99% (triple diversity)

Appendix II. Passive Satellite System Relay

Frequency	2,000 mc
Transmitter power	10 kw
Antenna size	60-foot parabolic
Antenna beamwidth	0.6 degree

Range	2,000 miles
Satellite height	460 miles
Number of satellites	15
Maximum slant distance	2,000 miles
Transmitted power	+ 40 dbw
Transmitter antenna gain	+ 50 db
Receiver antenna gain	+ 50
Free-space loss	− 338 db
Line losses	− 4 db
Echo-area gain	+ 62 db
Median received power	− 140 dbw
Noise figure	+ 9 db
Noise power per cycle	− 204 dbw
Receiver bandwidth $(B=3$ kc)	+ 35 db
Noise power	− 160 dbw
Median C/N	20 db
Minimum usable C/N	15 db
Fading allowance	5 db
Reliability	80% (no diversity) 99% (triple diversity)

References

1. American Ephemeris and Nautical Almanac (book). U. S. Government Printing Office, Washington, D. C., published annually.

2. The Air Almanac (book). U. S. Government Printing Office, published three times annually.

3. Navigation and Nautical Astronomy (book), B. Dutton. U. S. Naval Institute, Annapolis, Md., tenth edition, 1951.

4. Reference Data for Radio Engineers (book). International Telephone and Telegraph Corporation, New York, N. Y., 1956.

5. Radar Echoes from the Moon, J. Mofenson. *Electronics*, New York, N. Y., Apr. 1946, pp. 92–98.

6. Reflection of Microwaves from the Moon, Z. Bay. *Hungaria Acta Physica*, Budapest, Hungary, vol. 1, 1946, pp. 1–22.

7. Moon Echoes and Penetration of the Ionosphere, E. J. Kerr, et al. *Nature*, Washington, D. C., Feb. 26, 1949, pp. 310–13.

8. Moon Echoes and Transmission Through the Ionosphere, E. J. Kerr, et al. *Proceedings*, Institute of Radio Engineers, New York, N. Y., Mar. 1951, pp. 230–42.

9. Radio Echoes from the Moon, I. C. Browne, et al. *Proceedings*, Physical Society, London, England, Sept. 1956, pp. 901–20.

10. Detection of Radio Signals Reflected from the Moon, J. H. Dewitt, E. K. Stodola. *Proceedings*, Institute of Radio Engineers, 1949, pp. 229–42.

11. A UHF Moon-Relay, P. G. Sulzer, et al. *Ibid.*, Mar. 1952, p. 361.

12. UHF Moon Reflections, S. J. Fricker, et al. *Nature*, London, England, Nov. 22, 1958, pp. 1438–39 (digest).

13. Maximum Range of a Radar Set, K. A. Norton, A. C. Omberg. *Proceedings*, Institute of Radio Engineers, 1947, pp. 4–24.

14. Considerations of Moon-Relay Communication, D. D. Grieg, et al. *Ibid.*, May 1948, pp. 652–63.

15. The Influence of Radar Reflection Characteristics of the Moon on Specifications for Earth-Moon-Earth Communication Systems, T. B. A. Senior, et al. *Wescon Convention Record*, Institute of Radio Engineers, pt. 1, 1958, pp. 197–201.

16. Lunar Radio Echoes, J. H. Trexler. *Proceedings*, Institute of Radio Engineers, Jan. 1958, pp. 286–92.

17. Radar Echoes from the Moon at a Wavelength of 10 cm, B. S. Yaplee, et al. *Ibid.*, pp. 293–97.

18. Radio Astronomy (book), J. L. Pawsey, R. N. Bracewell. Oxford University Press, London, England, 1955, p. 301.

19. Basic Analysis on Controlled Carrier Operation of Tropospheric Scatter Communication Systems, L. P. Yeh. *Convention Record*, Institute of Radio Engineers, pt. 8, 1958, pp. 261–83.

20. Introduction to Space Communication System Design, L. P. Yeh. *Electronic Industries*, Philadelphia, Pa., Feb.–Mar. 1959, pp. 54–58 and 94–99.

21. The Physical and Chemical Fundamentals of Satellite Flight—Part II, P. F. Winterwitz. *Journal of Astronautics*, New York, N. Y., vol. 3, nos. 3–4, Autumn-Winter, 1956, p. 65.

22. Satellite Ascent Guidance Requirements, J. Jensen. *Ibid.*, no. 1, Spring, 1956, p. 1.

23. Problem of Launching an Earth Satellite, M. Summerfield. *Astronautics*, New York, N. Y., Nov. 1957, p. 18.

24. Lifetimes of Satellites from Near-Circular and Elliptic Orbits, N. V. Petersen. *Jet Propulsion*, New York, N. Y., May 1956, pp. 341–46.

25. Thermal Control in a Space Vehicle, P. E. Sandorff, J. S. Prigge. *Journal of Astronautics*, Spring, 1956, p. 4.

26. Skin Temperature of a Satellite, C. M. Schmidt, A. J. Hanawalt. *Jet Propulsion*, Oct. 1957, p. 1079.

27. *Aviation Week*, New York, N. Y., Mar. 24, 1958, p. 19.

Synthesis of a Resonant Transfer Filter as Applied to a Time Division Multiplex System

P. J. MAY
NONMEMBER AIEE

T. M. STUMP
NONMEMBER AIEE

IN A PREVIOUS PAPER,[1] Perkins describes a time division multiplex transmission path, developed for use in an electronic switchboard.[2] This transmission path incorporates the use of a "resonant transfer" filter, the operation of which was first described in reference 3. This type of filter and the associated theory have since been described by several authors.[4,5] The purpose of this paper is to present a method used to synthesize an economical network whose characteristics will allow it to operate efficiently in this system. The first sections cover the general requirements on the network and the special conditions imposed to effect a proper operation of the networks in a resonant transfer system. The general method of synthesis is then presented which involves both the steady-state and transient responses. Since the actual synthesis by use of well-known procedures of insertion parameter theory are familiar to most filter designers, the details of the synthesis are omitted and only the general procedure is given.

In Figs. 1, 2, and 3 is shown the transmission characteristics of a typical line circuit employing this type of filter design. Specifically shown are insertion loss, side-band suppression, envelope delay, and impedance as functions of frequency.

Filter Specifications

Referring to Fig. 4 where networks N_1 and N_2 are low-pass filters the following attenuation requirements must be met:

1. The network N_1 must attenuate frequencies above one half the switching frequency f_s. Because of the modulation process of the switch, input frequencies greater than $f_s/2$ can be translated into the audio band. As an example, if $f_s=12{,}500$ cps (cycles per second) and the generator signal is 9,000 cps, a frequency component of 3,500 cps is developed at the input to network N_2. Therefore, unless 9,000 cps is attenuated by N_1 it will cause a 3,500-cps signal to appear at the output.
2. The network N_2 is a smoothing filter which must attenuate the side bands about f_s developed by the modulation process. The exact attenuation specifications placed on N_1 and N_2 depend on the performance required of the system.

In addition to attenuation, specifications may be placed on the input and output impedance of the system and upon the envelope delay distortion.

Paper 60-822, recommended by the AIEE Communication Switching System Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting, Atlantic City, N. J., June 19–24, 1960. Manuscript submitted March 21, 1960; made available for printing April 19, 1960.

P. J. May and T. M. Stump are with the Stromberg-Carlson Division of the General Dynamics Corporation, Rochester, N. Y.

Credit is due to Heinz Kammin, who designed filters for this application, and to G. Sager, without whose thorough knowledge of magnetics the practical realization of these filters would have been impossible. Also appreciated is the assistance of various members of the Computations Group of Stromberg-Carlson Company in the solution of the inverse Laplace transforms on an electronic computer.

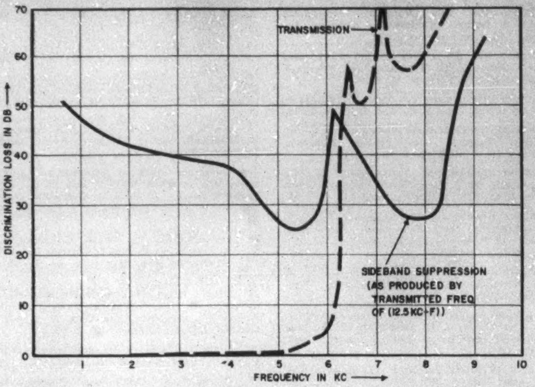


Fig. 1. **Insertion loss and side-band suppression of typical line circuit employing resonant transfer filters**

Conditions for Synthesizing Resonant Transfer Filters

Because of the periodically operating switch between the filters, a straightforward steady-state filter design will not necessarily give the required attenuation, phase, and impedance.

The following treatment will first state the design conditions imposed on the low-pass filter networks N_1 and N_2 of Fig. 4 and then show why these conditions are necessary. The three main conditions are:

1. The capacitors C_S and C_R must have a discharge curve which rings through zero at a period T equal to the sampling period. This type of curve is shown in Fig. 5.
2. The networks N_1 and N_2 are filters designed to work into an open-circuit termination, with capacitors C_S and C_R being at the open-circuit side.
3. For an impedance match

$$C_S = C_R = \frac{T}{2R} \tag{1}$$

The equality $C_S = C_R$ also assures a complete transfer of charge from network N_1 to N_2, provided L is adjusted so that

$$\tau = \pi\sqrt{LC}$$

where $C = C_S = C_R$ and τ is the switch closure time. The assumption that $\tau << t$ is also made.

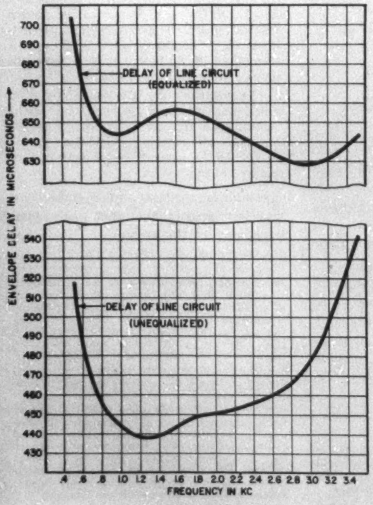
This treatment also supposes that series isolating coils L_1 and L_2 are incorporated in the filter designs, as shown in Fig. 4. This is done by requiring a full k section[6] performance in the design of the filter. Apart from the simplicity of analysis, it was found that the actual design was simplified by requiring this configuration. Otherwise, in requiring $C_S = C_R = T/2R$, the values of C_S and C_R would be equivalent input capacities, and would be difficult to control.

Theory of Operation

The operation of the switch is such that the voltage appearing across C_S before the switch closes appears across C_R when the switch opens again and, likewise, the voltage across C_R appears across C_S; see Fig. 6. Since capacitor C_R discharges its voltage in a period T and rings through zero in a period T, the voltage E_R is zero at each instant before the switch is closed. Therefore, no charge is transferred back to C_S, and hence, at each instant after the switch opens, E_S is zero.

The network N_1 may then be considered as driven by current pulses from the switch side, the amplitude of these pulses being of the exact value needed to bring E_S to zero at the end of each pulse.

The voltage E_S then becomes the sum of two voltages; the steady-state open-circuit voltage produced by the generator E_g, and the voltage produced by the



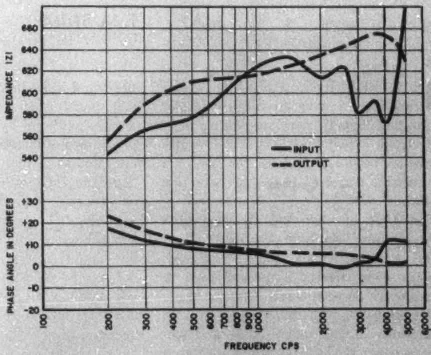


Fig. 2 (left). **Unequalized and equalized delay of line circuit**

Fig. 3 (above). **Input, output impedance and phase angle of line circuit**

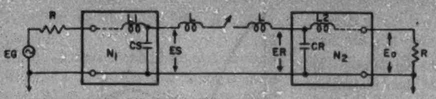


Fig. 4. Resonant transfer system

current pulses (superposition theorem). These voltages and their resultant are shown in Fig. 7.

The resultant voltage E_S at each instant prior to the switch closing, is at the exact value of the open-circuit voltage operating under steady-state conditions. This is true because it was required that capacitor C_S have a discharge curve which rings through zero in the period T.

The current pulses driving N_1 vary in amplitude following the steady-state open-circuit voltage of network N_1. These same current pulses also drive N_2 with an opposite polarity.

Since N_1 is designed as an open-circuit filter, the voltage E_{Soc}, and thus the amplitude of the current pulses as a function of frequency, can be controlled by the steady state design of N_1. In the pass band the amplitude of the current pulses varies as the ripple built into N_1. This is the reason for designing N_1 as an open-circuit filter. The network N_2 is effectively being driven, as shown in Fig. 8.

The current I consists of a pulse train modulated by a signal of the frequency of E_g, the envelope of which varies with frequency as the ripple of network N_1.

Examining the transmission of the fundamental frequency component, N_2 is operating from a constant current source into a finite load impedance. Since a current generator has an infinite impedance, and N_2 is designed to work into an infinite impedance on the C_R side, the transmission will be the same as the steady-state transmission of N_2. Thus, in the pass band the voltage E_0 will vary with frequency as the sum of the ripple of N_1 and N_2.

The filter N_2 must have a sufficient rejection to the other components of the current I shown in Fig. 8. The amount of rejection depends on the system specifications.

Impedance Matching

The current pulses composing I in Fig. 8 can be easily handled mathematically if the current consists of flat-topped pulses. Then the conventional pulse amplitude modulation equations may be used.[7] The current pulses of this system are actually half sine waves. In this analysis a flat-topped pulse of the same area will be used.

The equation for an individual current pulse is

$$I=\frac{E_{S-}}{2}\sqrt{\frac{C}{L}}\sin\frac{t}{\sqrt{LC}} \qquad (2)$$

where $t=0$ at the start of the pulse, $t=\tau=\pi\sqrt{LC}$ at the end of the pulse, and $E_{S-}=$voltage E_S at the instant the switch is closed. The area of this pulse is

$$\text{Area}=\int_0^{\pi\sqrt{LC}}\frac{E_{S-}}{2}\sqrt{\frac{C}{L}}\sin\frac{t}{\sqrt{LC}}\,dt = E_{S-}C \qquad (3)$$

A flat-topped pulse would have a width equal to τ and an amplitude of

$$A=E_{S-}\frac{C}{\tau} \qquad (4)$$

The envelope of the pulse train will be $Av=E_{Soc}C/\tau$, since the value of E_S follows the open-circuit voltage E_{Soc}. Av is now equal to the amplitude of the modulated signal in a pulse amplitude modulated system.

The modulating frequency component[7] is

$$I_{\text{fund}}=\frac{2E_{Soc}C}{\omega T\tau}\sin\frac{\omega\tau}{2} \qquad (5)$$

since

$$\tau\ll T \qquad (6)$$

$$T>\frac{1}{2f} \qquad (7)$$

Therefore

$$\sin\frac{\omega\tau}{2}\cong\frac{\omega\tau}{2} \qquad (8)$$

$$I_{\text{fund}}\cong E_{Soc}\frac{C}{T} \qquad (9)$$

For maximum power transfer N_2 would give no attenuation and the current would flow through the load resistance. The output voltage would then be

$$E_0=I_{\text{fund}}R \qquad (10)$$

and for maximum power transfer

$$E_0=I_{\text{fund}}R=\frac{E_g}{2} \qquad (11)$$

For maximum power transfer it must be assumed that N_1 gives no attenuation, so that

$$E_{Soc}=E_g \qquad (12)$$

Therefore

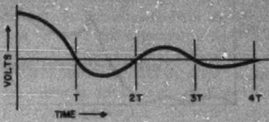


Fig. 5. Discharge curve of transfer capacitors C_s and C_r

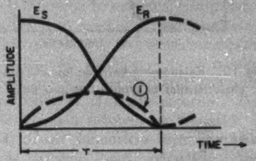


Fig. 6. Voltage versus time relationships during energy transfer in switch closure time

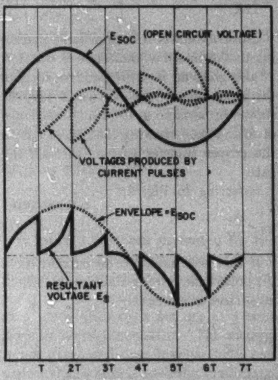


Fig. 7. Voltages produced across capacitor C_s

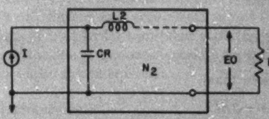


Fig. 8. Equivalent circuit of N_2

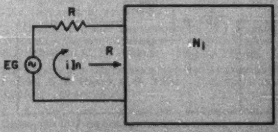


Fig. 9. Equivalent input circuit when impedances are matched

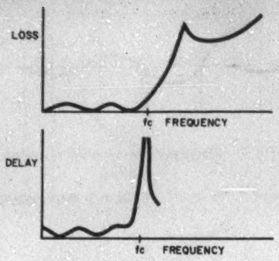


Fig. 10. Relationship between loss and delay characteristics of typical low-pass filter

$$I_{\text{fund}}R=\frac{E_{goc}}{2} \tag{13}$$

$$\frac{E_{goc}CR}{T}=\frac{E_{goc}}{2} \tag{14}$$

$$C=\frac{T}{2R} \tag{15}$$

This is a condition for impedance matching to obtain a maximum power transfer. It also gives an input impedance of R to the input of network N_1, which may be shown by assuming an input impedance of R and then showing that this leads to the proper transfer current through the gate.

Referring to Fig. 9,

$$i_{IN}=\frac{E_g}{2R} \tag{16}$$

If this is a fundamental current, then the pulse train flowing through the switch must have an envelope:

$$i_{\text{env}}=\frac{E_g}{2R}\,\frac{T}{\tau} \tag{17}$$

since

$$C=\frac{T}{2R} \tag{18}$$

Then

$$i_{\text{env}}=\frac{E_gC}{\tau} \tag{19}$$

Note that

$$\tau=\pi\sqrt{CL} \tag{20}$$

Then

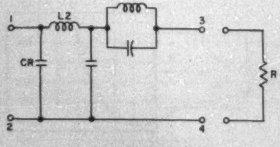


Fig. 11. Typical filter circuit for resonant transfer application

$$i_{\text{env}}=\frac{E_g}{\pi}\sqrt{\frac{C}{L}} \tag{21}$$

Finally, note that the actual pulses are half sine waves rather than rectangular pulses. Thus, the factor $\pi/2$ must be applied to obtain the peak amplitude of the envelope:

$$i_{\text{env}}=\frac{E_g}{2}\sqrt{\frac{C}{L}}=\frac{E_{goc}}{2}\sqrt{\frac{C}{L}} \tag{22}$$

This is identical to the solution of the resonant transfer circuit, which gives pulses of the same peak amplitude.

Filter Synthesis

The synthesis of this type of filter is not completely analytic nor exact, per se. It is rather an approximation to meet the specifications outlined above.

In view of the transient discharge and charge of capacitors in the resonant circuit which causes the transfer of energy from a send to a receive mechanism, one factor to be noted is that a type of isolation is required after the receive capacitor which causes all the charge to be stored there. The isolation chosen was an inductor which then becomes part of the filter design.

The filter could, of course, be designed with the use of the number of k sections[6] that would be necessary for the discrimination dictated by carrier and sideband suppression. In the interests of economy of elements, a design incorporating peaks of theoretically infinite attenuation at finite frequencies (m sections)[6] was chosen. This would give proper discrimination for the fewest number of elements.

Because of the isolation consideration mentioned and in view of the use of m sections, a nonmaximum species[8] filter was necessary. In the interests of economy, the insertion parameter approach to the filter design was chosen. This would give the optimum response for a given number of reactances of a given quality. Image parameter designs were considered as well, but the discharge response of these filters was found to be inadequate for this particular application. The insertion parameter design also gave another degree of freedom due to the choice of ripple. The terms "k section" and "m section" are borrowed from image parameter theory to describe the circuit, but not necessarily its operation. It is clear from the considerations in the previous section that the design had to be an open-circuit type. If the discharge curve is correct, the attenuation and phase of the filter in its circuit operation would approach quite closely the steady-state response, and could therefore be anticipated. This makes the design of the filter basically a steady-state problem.

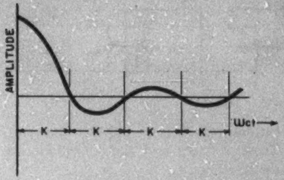


Fig. 12. Discharge of receive capacitor C_r through filter into load

Since the cutoff of the filter is determined primarily by the capacitor discharge transient, it was convenient to have other parameters to allow some freedom in the choice of the circuit. With a given circuit configuration there were three aspects to the synthesis which allowed the freedom desired (assuming insertion parameter design). These were value of ripple, pole location, and order of pole extraction.

As the title of this article indicates, the filter was applied to a time division multiplex system. From the simplest of considerations, then, the filter was to pass the audio of the transmission path, and reject the carrier (determined by the sampling rate) and the side bands. Once the carrier is fixed, many designs are possible which will give the proper discrimination to carrier and side bands.

Usually, extenuating circumstances force the design to a specific approach. In this particular application, it was the linearity of phase in the pass band of the filter which dictated the cutoff. As is well known, with the advent of high-speed data transmission, linearity of phase with its resultant low envelope delay distortion is mandatory on transmission paths. There are two requirements on a transmission path, as far as delay is concerned, for the faithful reproduction of pulse transmission: 1. linearity of phase resulting in low envelope delay distortion, and 2. zero or $n\,\pi$ zero phase intercept.

The second point is obviated with the use of modulation. That is, the transmission will be faithful if a modulated signal is transmitted through a nonzero (including $n\pi$) phase intercept system, and if (assuming no attenuation distortion) the phase-versus-frequency characteristic is linear over the frequency band of transmission. For base-band transmission, the zero intercept must be considered. Because of the linear phase consideration, the cutoff of the filter was kept as high as possible.

Fig. 10 depicts the phase and envelope delay of a typical low-pass filter configuration. The slope of the phase curve rapidly increases in the vicinity of cutoff so that an attempt must be made to keep the cutoff as far removed as possible from the frequencies over which the phase is to remain linear. All pass phase equalizer sections with complementary phase characteristics can be added to the transmission path to increase the linearity of phase. However, an effort to reduce the phase distortion in the first place makes the equalization problem much easier and also keeps the absolute level of envelope delay down to a minimum. This last factor becomes important where synchronization based on delay of a system becomes important.

Certain design approaches yield much more linear phase in the pass band and through the cutoff. However, these designs have two drawbacks:

1. An increasing number of loss-versus-frequency characteristics, which is undesirable in practical transmission paths.
2. The large number of sections required for good rejection to a carrier and side bands not too remote from cutoff, leading to poor economy of design.

These considerations will yield an approximation to the circuit configuration. Assignment of circuit parameters will yield a normalized filter circuit. The normalized filter circuit is the result of a synthesis procedure for a filter using insertion parameter theory developed by Cocci,[8] Darlington,[9] Cauer,[10] and others. The Darlington reference filter was found to be quite adaptable to this design due to the arbitrariness of the possible pole location. A Cocci 5–1, 7–1, or 7–2 could also be used with success. All of these designs, as mentioned previously, must be of open-circuit variety where the open-circuit end faces the gate for sampling. This circuit will have a normalized transient response in terms of the cutoff. If this response is correct and yields a cutoff which is placed so as to allow the filter to reject the carrier and side bands, and allows room for phase linearity, the network becomes theoretically feasible.

An important consideration in the design of a resonant transfer filter is the determining of the cutoff as a result of the capacitor discharge transient. The rate of sampling in the time division system is determined by the carrier frequency used, which also determines the time which can be subdivided into the desired number of channels. The sample time is then determined by the duration of time through which the switch is closed. We shall call the sampling rate T and the sampling time τ. The duty cycle is then given by τ/T. As an example, if a 12.5 kc carrier is used, the highway time available would be 80-μsec (microseconds). This time could be subdivided into 32 time slots, each 2.5 μsec in duration. This could allow a 1.25-μsec sampling time and a 1.25-μsec guard space, which would allow 32 possible conversations on one highway. Since each time slot is sampled every 80 μsec, the voltage on the receive capacitor must be zero just prior to the closing of the switch. This is why the discharge of this capacitor must be carefully controlled. The analysis of a typical case would be as follows.

Assume a given filter configuration which must meet the transmission requirements on the system. Such a filter might look like that shown in Fig. 11.

Assume a charge of 1 on the capacitor C_R. Write the Laplace transform of this network as a function of the complex frequency variable, $S=\sigma+jw$. With the use of the inverse transform the time function of this network may be determined. If one then considers the discharge of this capacitor across terminals 1 and 2 through the filter into the load, Fig. 11, the curve might be like that in Fig. 12.

The ideal case is where zero crossings of the discharge occur as integral multiples of Wct. This would assure a zero value of charge on the capacitor each time a new sample is made. The transfer of energy from send to receive would be lossless, assuming that other matching conditions are met. The first zero crossing as a function of Wct could then make it possible to arrive at a value of Wc which corresponds to the normalizing cutoff frequency of the filter. The use of this frequency to unnormalize the filter would assure the designer of the proper transient response, if $Wc=K/T$, where K is the normalized time of zero crossing. As a matter of fact, any of the zero crossings could be used to normalize the filter. Unfortunately, a perfect integral multiple zero crossing of Wct is not practicably possible. Variation of the filter parameters will yield a good approximation of this goal.

Once the value of Wc is determined, it is compared to the requirements on the filter due to the transmission specifications. If the cutoff is in the proper location, the filter is assumed to be a good approximation to the design required.

The last requirement on the filter is that its discharge capacitor be of the proper value if an impedance match is desired, as has already been described. The value of this capacitor is $C=T/2R$, where T is the sampling interval and R is the impedance level of the filter, i.e., the send filter is driven from an R-ohm impedance generator and the receive filter is terminated in an R-ohm load. If the capacitor value is different from the prescribed value, a mismatch loss will result, as described earlier.[5]

It is possible by empirical manipulation to arrive at a set of circuit parameters which will yield a result which satisfies the simultaneous fulfillment of the requirements on both the discharge curve and the value of discharge capacity. This manipulation involves the variation of the circuit parameters such as ripple, pole location, and extraction. Once this is determined, the resonance of the transfer circuit is fixed by the sampling time. As shown in reference 1, the capacitor voltage-versus-time plot resembles that of Fig. 6.

If, then, in the example above, resonance of the transfer circuit is made to be 400 kc, the transfer of energy will be accomplished in the 1.25 μsec in which the switch is closed. The resonance could be arranged as shown in Fig. 4, where the series resonance, or C_S, C_R, and $2L$, is at 400 kc.

Conclusions

A method has been presented for the synthesis of a filter which operates from a finite source impedance into a load which is actually an electronic switch operating at a given duty cycle for use in a time division multiplex system. The method is not exact and does involve some empirical trial which is basically mathematical. Once the requisites for the efficient operation outlined are met, one may feel confident that the design will operate effectively. The resonant transfer system is an efficient and economical means of energy transfer for transmission purposes, and the design approach is mainly dependent upon this filter circuit.

References

1. TRANSMISSION ASPECTS OF AN ELECTRONIC SWITCHBOARD EMPLOYING TIME-DIVISION MULTIPLEXING, J. C. Perkins. *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 78, 1959 (Jan. 1960 section), pp. 949–53.

2. CONSIDERATIONS PERTAINING TO THE DESIGN OF AN ELECTRONIC TELEPHONE SWITCHING SYSTEM, B. Brightman, M. P. Tubinis. *Ibid.*, pp. 777–86.

3. MEANS FOR DETECTING AND/OR GENERATING PULSES, H. B. Haard, C. G. Svala. *U.S. Patent no. 2,718,621*, issued 1955.

4. A NETWORK CONTAINING A PERIODICALLY OPERATED SWITCH SOLVED BY SUCCESSIVE APPROXIMATIONS, C. A. Desoer. *The Bell System Technical Journal*, New York, N. Y., vol. 36, no. 6, 1957, p. 1403.

5. AN EFFICIENT ELECTRONIC SWITCH—THE

Bothway Gate, J. A. T. French, D. J. Harding. *Post Office Electrical Engineers' Journal*, London, England, vol. 52, pt. I, Apr. 1959, pp. 37–42.

6. Transmission Networks and Wave Filters (book), T. E. Shea. D. Van Nostrand Company, Inc., Princeton, N. J., pp. 221–60.

7. Reference Data for Radio Engineers. International Telephone and Telegraph Corporation, New York, N. Y., fourth edition, pp. 438–48.

8. Filters with a Minimum Number of Elements, G. Cocci. *Alta Frequenza*, Milan, Italy, vol. 11, Dec. 11, 1942, pp. 452–556.

9. Synthesis of Reactance 4-Poles, S. Darlington. *Journal of Mathematics and Physics*, Cambridge, Mass., vol. 18, Sept. 1939, pp. 257–353.

10. Synthesis of Linear Communication Networks (book). W. Cauer. McGraw-Hill Book Company, Inc., New York, N. Y., vol. I and vol. II, 1958.

11. Synthesis of Tchebycheff Parameter Symmetrical Filters, A. J. Grossman. *Proceedings*, Institute of Radio Engineers, New York, N. Y., Apr. 1957, p. 454.

12. The Network Synthesis on the Insertion Loss Basis, J. Zdunek. *Monograph no. 278R*, Institution of Electrical Engineers, London, England, Jan. 1958.

13. Transients in Linear Systems (book), M. F. Gardner, J. L. Barnes. John Wiley & Sons, Inc., New York, N. Y., vol. I, 1942.

An Electronic Teleprinter Distributor

N. N. BISWAS
ASSOCIATE MEMBER AIEE

H. RAKSHIT
NONMEMBER AIEE

ALTHOUGH MOST present-day teleprinters have mechanically operated cam-type distributors, which are quite satisfactory and reliable, a recent trend in the industry has been to develop electronic ones.[1,2] This paper describes an electronic distributor developed at the electrical engineering laboratory of the University of Roorkee, India. The distributor uses a ring counter consisting of cold cathode trigger tubes. The timing pulses are derived from the 50-cycle main supply. The 1.5-unit stop pulse is insured by a novel method of reversing the phase of the sinusoidal supply driving the pulse generator during the stop period of each signal train. Since no hot cathode tube is used at any stage of the distributor, it does not require any warm-up time. The distributor can be fed either directly from the keyboard or from a perforated tape. It can also be made to work from a 5- or 6-unit code by means of a code-selector switch.

The Distributor

Fig. 1 shows the block diagram of the electronic distributor. $C1$ to $C5$ are the five code contacts which are influenced either by the depression of a key bar or by the perforation of a type. These contacts are scanned by the ring counter in proper interval and sequence to transmit the mark or space as the case may be.

Ring Counter

The ring counter employs cold cathode thyratrons or trigger tubes connected in the usual 3-4 manner. To transmit a character, the distributor is required to send 7 elements: start, 5 code elements, and the stop element. At this stage let the stop element be of 2 units duration instead of the usual 1.5 unit. Fig. 2 shows that this will require 8 pulses and, therefore, 8 tubes. The sequence of operation is as follows. At the beginning of the transmission, tube T_0 is made to conduct. When the S pulse comes, tube T_S fires; when pulse 1 comes, tube T_1 fires. During the interval between the firing of tube T_S and tube T_1, the start impulse is transmitted. Then comes pulse 2 which fires tube T_2.

During this interval the first code element is transmitted. Similarly, by the time pulse M arrives, all the code elements have been transmitted. Pulse O is the last pulse for the character. This fires tube T_o, which remains conducting until the arrival of the next train of pulses for the next character. It may be seen that when two characters are transmitted one after another, as in tape transmission, the interval betwern the firing of T_o and that of T_S is 1 unit. This makes the duration of the stop element, which is the duration between the M pulse and S pulse, equal to 2 units.

It has been said that the timing pulses are derived from the main supply. When the main supply has a frequency of 50 cycles per second, the teleprinter will operate at a speed of 50 bauds. The pulse generator is so designed that, for reasonable variation of the supply voltage, there is no appreciable change in the amplitude of the output pulse. Any fluctuation of frequency will, however, affect the pulse repetition frequency. But it has been seen that if the frequency variation is limited to 4% a teleprinter receiver having a margin of 35% will function properly.

Send Relay

A high-speed polarized telegraph relay acts as the output or send relay of the distributor. The SPDT contact of the relay can be wired to provide either neutral or polar operation. Fig. 3. shows how, when tube T_1 of the ring counter conducts, the current through the send relay flows through either the 2-1 or 3-4 winding depending on the condition of the code contact $C1$. Current through the 2-1 winding makes the relay space, and current through the 3-4 winding makes the relay mark.

Before the transmission of the first code element, the start signal must be transmitted. For this reason tube T_S precedes tube T_1 in the ring counting cycle and circuit; see Fig. 4. Again, since the start signal is always a space pulse, tube T_S has been connected permanently to 1-2 winding of the send relay. For a similar reason, during the conducting period of tube T_M, the stop pulse, always marked, is transmitted. Tube T_M has been connected permanently to the 3-4 winding of the send relay immediately after tube T_5 or T_6. When the transmitter is idle, marking current should flow into the line. Hence, tube T_o, which continues to conduct during the idle period is also connected permanently to the 3-4 winding of the send relay; see Fig. 4.

Control Relay

It is clear from Fig. 2 that 8 pulses are needed to transmit a teleprinter signal for a character. The transmission of the fifth code element is over, however, when the seventh pulse comes. Therefore, up to this time it is essential that the code contacts, combination bars, and function bar are all in operation and undisturbed. This is achieved by keeping the key bar mechanically locked during this period of transmission. During this locked condition the operator is unable to depress any other key bar. As-

---

Paper 60-770, recommended by the AIEE Telegraph Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting, Atlantic City, N. J., June 19–24, 1960. Manuscript submitted February 10, 1960; made available for printing March 29, 1960.

N. N. Biswas is with the University of Roorkee, Roorkee, India, and H. Rakshit is with the Indian Institute of Technology, Kharagpur, India.

This paper is derived from a thesis submitted by N. N. Biswas to the Indian Institute of Technology, Kharagpur, India, in partial fulfillment of the requirement for a Ph.D. degree. The work was supervised by Professor H. Rakshit. The authors wish to acknowledge the help and encouragement of Dr. A. N. Khosla of the University of Roorkee, India.

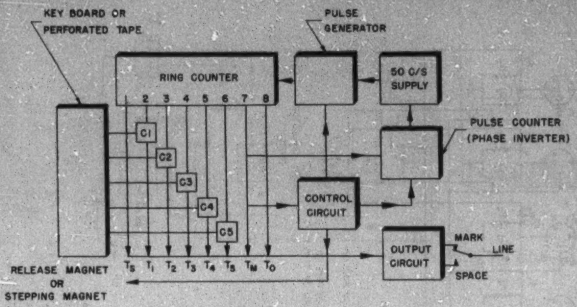

KEY BOARD OR PERFORATED TAPE
PULSE GENERATOR
RING COUNTER
1 2 3 4 5 6 7 8
50 C/S SUPPLY
PULSE COUNTER (PHASE INVERTER)
C1
C2
C3
C4
C5
CONTROL CIRCUIT
OUTPUT CIRCUIT
MARK
LINE
SPACE
RELEASE MAGNET OR STEPPING MAGNET
T_S T_1 T_2 T_3 T_4 T_5 T_M T_O

Fig. 1. Block diagram of electronic distributor

suming the characters are transmitted one after another without any loss of time, it is obvious that the stop signal is transmitted between the seventh and ninth pulses. During this period the keyboard should be released and a fresh combination set up. This is done by the control relay.

If the operator keeps the key depressed for a period longer than the transmission time of a signal, the transmitter should transmit the signal only once. Here also it is the control relay which prevents this undesirable repetition of the signal. Fig. 5 shows the circuit arrangement of the control relay along with the pulsing circuit. The control relay is also a high-speed polarized relay similar to the send relay, but has two changeover contacts. In Fig. 5 only the operating winding has been shown. A small biasing current always flows through the other winding, which is not shown. The contact *CR*1 is the contact of the control relay. *F*1 and *F*2 are two contacts actuated by the function bar and hence by the depression of the key.

If the operator keeps the key depressed for a period less than the transmission time, the control relay functions in the following manner. As soon as the key is depressed, the code contacts, *C*1 to *C*5 in Fig. 4, as well as the function contacts *F*1 and *F*2, are locked in the operating position mechanically. The pulsing circuit is grounded via *CR*1 and *F*1, starts generating the pulses, and transmission begins. At the end of transmission of the fifth and final code pulse, tube T_M fires and the control relay operates. Contact *CR*1 changes over and grounds the release magnet of the locking mechanism. The release magnet is very fast operating, and unlocks the locked contacts in less than 10 milliseconds. Contacts *F*1 and *F*2 come to normal position. *F*2 again grounds the pulsing circuit, and the next pulse fires tube T_o and extinguishes tube T_M. The control relay releases immediately, bringing contact *CR*1 to normal position. As a result, the ground of the pulsing circuit is removed and no further pulse is generated. Tube T_o continues to conduct, and the distributor is ready for transmission of the next signal.

On the other hand, when the key is kept depressed for a longer time than the transmission period, the code contacts and the function contacts operate and are locked in the usual way following the depressing of the key. At the end of the transmission of the fifth code element, tube T_M fires. The control relay operates, *CR*1 changes over, and the pulse generator is de-energized. Although at this instant the release magnet of the locking mechanism is energized the code contacts and the function contacts do not come to normal position as the combination and function bars remain operated by the depression of the key bar. The distributor, therefore, remains in this position as long as the key is depressed by the operator. Marking current continues to flow onto the line, since the tube T_M remains conducting. As soon as the key is released, all contacts come to normal position. Pulse generator is energized once again via *CR*1 and *F*2. The first pulse now generated fires tube T_o and extinguishes tube T_M. Consequently, the control relay releases. *CR*1 comes back to normal position, de-energizing the pulse generator. Tube T_o maintains marking current on the line, and the distributor is ready for sending the next code train.

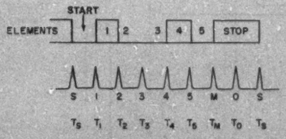


Fig. 2. Pulses and tubes required to transmit teleprinter signal

Run-out Key

In the mechanical teleprinters a special key is provided which, when depressed, sends the code of the character last transmitted over and over again. The key is known as the run-out key. Here the run-out key takes the form of a telephone key and is on the control board rather than on the keyboard. If repeated transmission of a particular character is desired, first the run-out key is operated; see Fig. 4. This operates the two contacts *RT*1 and *RT*2. Now the key for the particular character is depressed. The code contacts and the function contacts are locked. It may be seen that in spite of the operation of *CR*1, the pulse generator is working because of the ground extended by *RT*1, and the release magnet is not grounded because of *RT*2. Hence the transmission is repeated until the run-out key is brought back to normal position.

Stop Pulse

The principle by which a stop pulse of 1.5 units is obtained is shown in Fig. 6. As soon as tube T_M fires, the phase of the sinusoidal voltage driving the pulse generator is reversed. As a result the next pulse comes after 10 milliseconds and fires tube T_o. Thus, the interval between T_M and T_o is reduced to 1/2 unit instead of 1, and the stop pulse, which is the interval between the firings of tube T_M and tube T_S, now has a duration of 1.5 units. To carry out this principle two conditions are required: a means for indicating the time at which the supply is to be reversed, and a means for reversing the supply voltage at the appropriate time.

The control relay in conjunction with

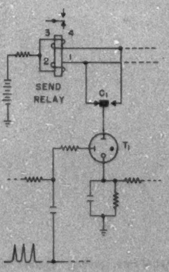
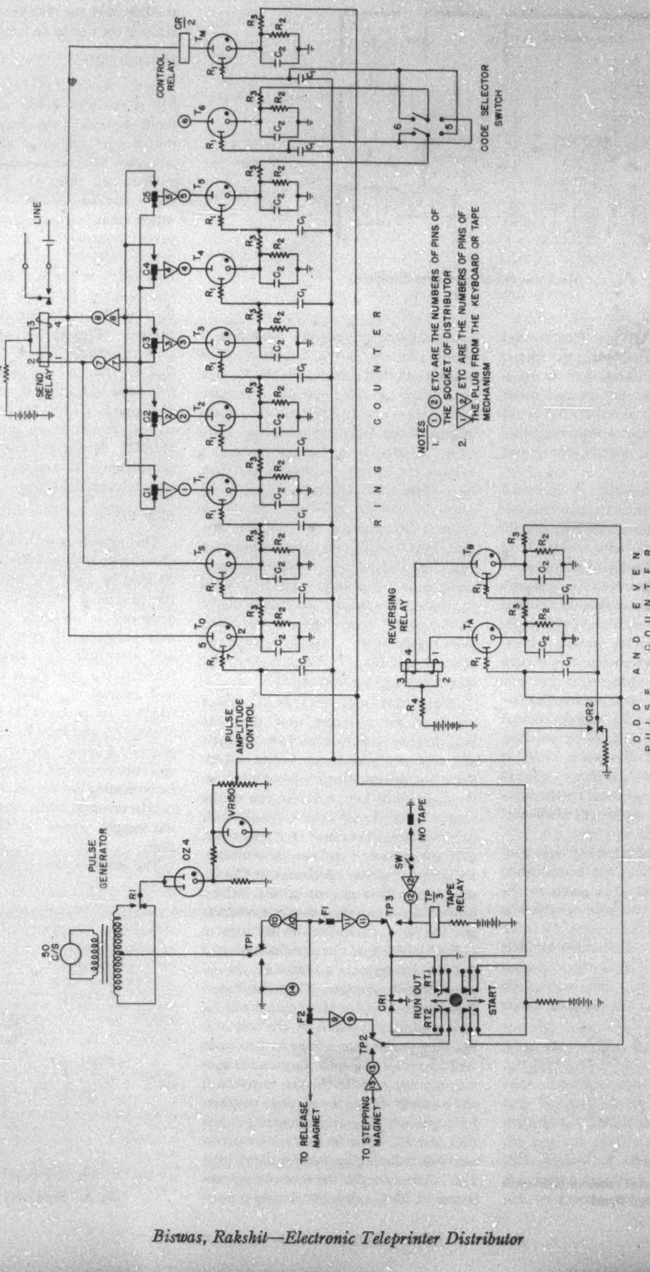


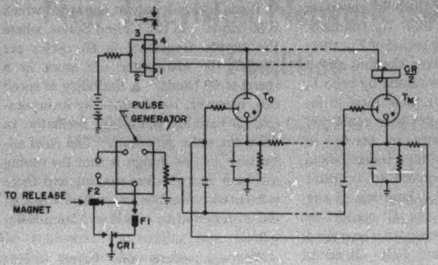
Fig. 3. Send relay circuit


50 C/S
PULSE GENERATOR
RI
OZ 4
VR150
PULSE AMPLITUDE CONTROL
TPI
TP2
TP3
F1
F2
TO RELEASE MAGNET
TO STEPPING MAGNET
SW
NO TAPE
TP 3
TAPE RELAY
CRI
RUN OUT
RT1
RT2
START
T_0
T_S
T_1
T_2
T_3
T_4
T_5
T_6
T_M
R_1
R_2
R_3
R_4
C_1
C_2
C1
C2
C3
C4
C5
SEND RELAY
LINE
CONTROL RELAY
CR 2
CODE SELECTOR SWITCH
5
6
R I N G   C O U N T E R
REVERSING RELAY
T_A
T_B
CR2
ODD AND EVEN PULSE COUNTER

NOTES

1. ① ② ETC ARE THE NUMBERS OF PINS OF THE SOCKET OF DISTRIBUTOR
2. ▽1 ▽2 ETC ARE THE NUMBERS OF PINS OF THE PLUG FROM THE KEYBOARD OR TAPE MECHANISM

Fig. 4. Complete circuit of electronic distributor



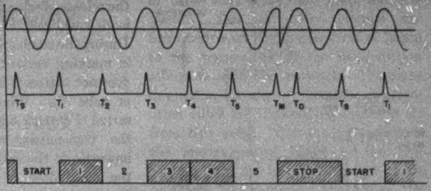


Fig. 5 (left). Control relay and associated circuit

Fig. 6 (above). Principle of insuring a stop pulse of 1.5 units

the odd and even pulse counter provides a convenient way to achieve these means. The circuitry of the odd and even pulse counter, seen in Fig. 7, is the same as that of the ring counter. The same trigger tubes with similar resistors and capacitor connections are used. The reversing relay is similar to the send relay. Its contact *R*1 reverses the phase of the sinusoidal supply driving the pulse generator; see Fig. 4.

When the control relay operates with the firing of tube T_M, contact *CR*2 seen in Fig. 7, changes over almost instantaneously, connecting the starter anodes of tube T_A and T_B to the output of the pulse generator. As a result, the same pulse which fires tube T_M makes these two tubes interchange their states. The next pulse now comes after 10 milliseconds. This pulse fires tube T_0 of the ring counter, but should not affect the states of the tubes T_A and T_B. This is accomplished by increasing the values of C_2's in the odd and even pulse counter. Because of the greater value of C_2, the charging time is increased and the voltage across R_3 which provides the positive bias to the starter anode does not rise sufficiently within 10 milliseconds. Consequently, this pulse fails to fire any tube of the odd and even pulse counter. Immediately after this pulse, *CR*2 comes back to normal; the tubes T_A and T_B interchange their states only when tube T_M fires once again. Thus the *R*1 contact of the reversing relay reverses the phase of the sinusoidal supply during the transmission of the stop pulse of each signal train. This insures a 1.5-unit stop pulse.

Fig. 8 shows the oscillograms of the sinusoidal supply, the pulses, and the output wave form when the distributor was made to transmit the code for letter D continuously with the help of the run-out key. The oscillograms were taken by feeding these three quantities simultaneously on 3 elements of a 6-element magnetic oscillograph.

START KEY

The ring counter and the odd and even pulse counter are not self-starting. That is, when power is switched on, none of the tubes in these circuits start conducting. A high voltage is applied manually by the start key shown in Fig. 4, to the starter anodes of tube T_0 and tube T_A, and then removed. These two tubes then start conducting and the distributor is ready for transmission.

TAPE TRANSMISSION

In tape transmission the distributor works continuously. During the transmission of 30-millisecond stop pulse the tape-feeding mechanism is energized. This moves the tape by one hole, presenting the code of the next character before the sensing needles. For this, an additional relay *TP* is employed. Relay *TP* is a conventional telephone relay with three contacts. This relay remains operated for the entire period of tape transmission. The circuit of this relay and its contacts are shown in Fig. 9.

In order that the distributor can work as a keyboard transmitter and as a tape transmitter, the 14 points as shown in Fig. 4 are wired into a multicontact socket. When the distributor works in conjunction with a keyboard, the code contacts, function contacts, and other points which are on the keyboard are connected to the appropriate points of the distributor by means of the plug from the keyboard. When the distributor is required to work from a tape mechanism the keyboard plug is disconnected and the plug from the tape mechanism is inserted; this brings in the code contacts and other points from the tape mechanism to the appropriate points of the distributor.

After the tape is inserted, relay *TP* operates when switch *SW* is closed. *TP*1 energizes the pulse generator by grounding it and transmission commences. At the

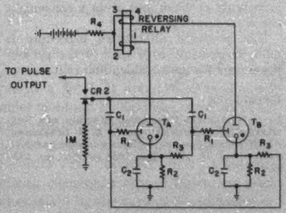


Fig. 7. Odd and even pulse counter

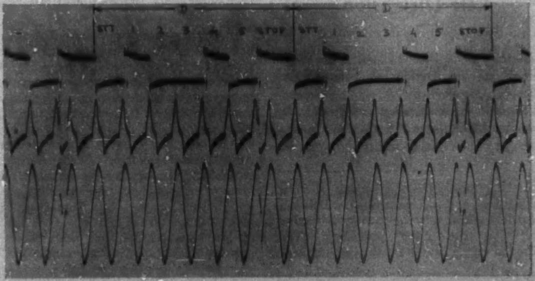


Fig. 8. Oscillograms showing phase reversal of sinusoidal supply, pulses, and output wave form

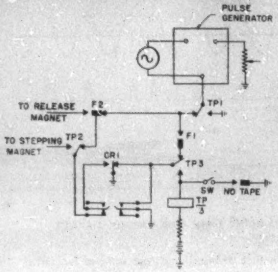


Fig. 9. Tape relay circuit

end of the transmission of the fifth element, tube T_M fires and the control relay operates. *CR*1 energizes the stepping magnet of the tape feed mechanism via *TP*2. The tape moves by one feed hole, and the code of the next character is presented before the sensing needles. This cycle is repeated for every character. At the end, when the entire tape has moved over, the switch NO TAPE releases, releasing relay *TP*. It may be seen that if relay *TP* releases at any other time than when the stop pulse is being transmitted, pulsing may stop when any of the tubes T_1 to T_5 is conducting.

Because of NO TAPE all the code contacts are in the spacing position; thus, the distributor may come to a stop with, say, tube T_3 conducting, and keeping spacing current over the line. Since this is not desirable, relay *TP* is prevented from releasing at any other time by grounding by *TP*3 and *CR*1. When tube T_M is conducting *CR*1 removes the ground and relay *TP* releases. Release lag of relay *TP* is less than 10 milliseconds. With the release of relay *TP* the pulse generator is de-energized, and tube T_M conducts. During the conducting period of T_M, there is marking current over the line, and it does not matter if it is caused by tube T_0 or tube T_M. When a new tape is inserted, *TP* relay operates once again and the transmission starts as usual. If, however, at the end of tape transmission, the distributor is required to transmit from the keyboard, the insertion of key board plug brings in the *F*2 contact of Fig. 9, and the pulse generator is energized via *CR*1, *TP*2, *F*2, and *TP*1. It sends out one pulse which fires tube T_0 and extinguishes tube T_M. The control relay releases, *CR*1 comes back to normal, and the pulse generator is de-energized.

FLEXIBILITY OF CODE INPUT

In some countries, as in India, the complexity of the language may be more than a 5-unit code can represent.[5] It is therefore desirable that the distributor work from a 6-unit code also. In the distributor that has been developed, this facility has been provided by simply having one more tube in the ring counter, as shown in Fig. 4. The circuit of this tube is essentially the same as that of any other tube of the ring counter. When the transmitter is working from a 5-unit code, the *DPDT* switch removes this tube from the ring counter. When the switch is in the 6-unit position, the tube is reinstated and the distributor then works from a 6-unit code input.

Fig. 4 gives the complete circuit of the electronic distributor.

Conclusions

The electronic distributor as described here has one important limitation, that is, it can transmit only at one speed which is 50 bauds. In the United States, where the supply frequency is 60 cycles per second, the distributor will work at a speed of 60 bauds. A flexibility of speed can, however, be obtained by incorporating a variable frequency oscillator to drive the pulse generator. The chief advantage of this design is that the timing oscillator remains free running and there is no need to double its frequency during the stop period to obtain a 1.5-unit stop pulse.

Although there are better methods of producing sharp pulses using vacuum tubes, a gas tube has been used in the distributor so that no warm-up time is necessary. A peaking transformer can also be used conveniently as a pulse generator. For a similar reason the power supply should also use a gas or a metal rectifier. The ring counter can be made more compact by using transistors or dekatron tubes.[6]

References

1. 5-UNIT TELEPRINTER AUTONUMBERING TRANSMITTER—TAA-1. *Publication REB.107*, Automatic Telephone and Electric Company, Ltd., London England, p. 4.

2. AIRBORNE TELETYPEWRITER AN/AGC-1, R. A. Michals. *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 77, May 1958, pp. 234–39.

3. THE DESIGN OF SWITCHING CIRCUITS (book). W. Keister, A. E. Ritche, S. H. Washburn. D. Van Nostrand Company, Inc., Princeton, N. J., 1957, p. 265.

4. INDUSTRIAL ELECTRONICS, R. Kretzmann. *Philips Technical Review*, Eindhoven, Netherlands, 1953, pp. 73–74.

5. TELEPRINTER CODE FOR HINDI ALPHABETS, N. N. Biswas. *Journal*, Institution of Telecommunication Engineers, New Delhi, India, vol. 3, June, 1957, pp. 235–41.

6. THE DEKATRON—A NEW COLD CATHODE COUNTING TUBE, R. C. Bacon, J. R. Pollard. *Electronic Engineering*, London, England, vol. 22, 1950, pp. 173–77.

◆

Late Discussion

Discussion of "The Numerical Graphical Method in the Design of Multiterminal Switching Circuits," by A. H. Scheinman, published in Communication and Electronics, November 1959, pages 515–19.

C. A. Stapleton (The University of New South Wales, Sydney, Australia): This is a useful extension of the author's earlier paper[1] on the same procedure. The method is very suitable for the "noncareer" switching-circuit designer who can rapidly become proficient in its use.

In neither paper does the author discuss how to weight the inputs, i.e., in what order should they be removed in the design. It is my experience that this order of removal is important in single-output problems with many inputs and I suspect that this would carry over to the method *1* of this paper; the author agrees that it affects method *2*. An alternative method of representing the switching function, by means of a transmission matrix, has been described in an unpublished paper by Stapleton. This procedure has all the features of the numerical-graphical method but adds some important advantages. For example, it enables the selection of the removal order which gives efficient results, particularly when a number of input combinations are don't-care items and permits a change in removal order after the reduction process has begun. It reduces the paper work in a large-scale problem and gives a convenient check on solutions.

The proposed alternative procedure is quite compatible with that of the present paper and is a useful addendum to it.

REFERENCE

1. A NUMERICAL-GRAPHICAL METHOD FOR SYNTHESIZING SWITCHING CIRCUITS, A. H. Scheinman. *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 76, 1957 (Jan. 1958 section), pp. 688–89.

New Television Network Switching Facilities

A. E. BACHELET
NONMEMBER AIEE

C. A. COLLINS
NONMEMBER AIEE

E. R. TAYLOR
MEMBER AIEE

"WE NOW return you to our studio in New York." At this cue from the announcer the smiling face of the weather man in Chicago disappears from the screen of your television set and is replaced by that of a well-known news commentator. This change in the originating point of a television program is a familiar phenomenon to viewers and similar "switches" are performed in each television network many times each day. However, back of the ability of the various television networks to originate or receive programs almost anywhere in the United States lies a large web of intercity and local audio and video transmission and switching facilities leased by the Bell Telephone Company system to the television networks. These facilities are constantly being expanded and improved and it is the purpose of this paper to describe briefly the video switching facilities previously used and to present a more detailed description of new video switching facilities which are now being put into use.

l. 79
1
60

Intercity and Local Facilities

At the present time, three major networks, the National Broadcasting Company (NBC), American Broadcasting Company (ABC), and Columbia Broadcasting System (CBS), carry most of the television programs of the country. Programs of these networks largely originate in New York, Los Angeles, Chicago, or Washington, D.C.; but they may originate and be introduced into a network at almost any other point. The audio portion of each program is carried over facilities, built up largely or use in radiobroadcasting, which are separate from those furnished for the video portion. Fig. 1 shows a map of intercity video television facilities as they existed on January 1, 1960. These facilities comprised a total of about 85,000 channel miles of which 73,000 miles were on TD-2 (transmitter-distributor) radio relay and 13,000 miles in coaxial cable. These facilities are divided into numerous circuit elements many of which are used alternately by the various broadcasters or closed-circuit networks. This alternate use requires frequent switching of intercity circuit units as the needs of the broadcasters vary from hour to hour. In certain areas, where a number of circuits in each direction are available, the amount of switching is reduced by the use of a circuit arrangement which is termed a "round robin." As an illustration of this arrangement, the NBC television network is furnished with facilities which go west from New York through Philadelphia, Baltimore, Washington, Pittsburgh, Columbus, Dayton, and Indianapolis to Chicago and return to New York via east-bound facilities through Cleveland, Buffalo, and Albany. CBS is furnished with similar facilities which go through the same cities, but in the opposite direction. Other round robins involving other cities are gradually coming into existence. A network may wish to make any one of these cities the transmitting point for a particular program, with all others receiving; or it may wish to change transmitting points during a program as in the case cited in the opening paragraph. The television facilities of Fig. 1 must, therefore, include suitable switching gear not only to interconnect intercity circuit units but also to connect the Bell System intercity network facilities to local transmitting and receiving video circuits extending to the various studios or stations affiliated with the networks within each city. This switching gear is located in television operating centers (TOC) in most large towns and cities and also at the radio towers or coaxial terminals that serve each city.

Typical video facilities which might be furnished by the Bell System to one of the television networks within a particular city are illustrated in Fig. 2. The local video facilities furnished to a network radiate from a master-control board owned and operated by the broadcaster. To this point are brought one or more video circuits from each program originating point within the city. Some of these, from the local studios of the network, are used regularly for program origination. Others, from various local pickup points, such as the White House in Washington or Madison Square Garden in New York City, are used less frequently. One or more circuits are furnished from the master-control board to the transmitter of the local station affiliated with the network. Programs originating locally or brought in from other cities are transmitted over the air from this point. Circuits to and from the master-control board to the Bell System TOC are furnished for use on all programs which involve the use of the intercity facilities assigned to the network. The TOC has interconnecting facilities to the TD-2 radio relay tower or coaxial terminal serving the particular city.

Switching Facilities

Fig. 1 indicates that a very large percentage of the intercity routes have channels equipped for color transmission and various improvements have been introduced which have made this possible. Local video channels used to interconnect the television broadcaster's various locations within a city and to connect the broadcaster to the Bell System intercity network are being improved by the introduction of the new *A2A* video transmission system. Comparable improvements have also been made in the video switching facilities to be provided in TOC's such as the one shown in Fig. 2. The video switching arrangements provided in such a center permit switching a number of incoming intercity facilities or facilities from local network master-control boards to any one or more of a number of outgoing circuits which may also be intercity facilities or facilities to a local master-control board. The video switch originally used for this purpose is shown in Fig. 3. It employs multibutton key sets to bridge high-impedance video output circuits to incoming circuits. This type of switch required special equalization and careful limiting of the number of outputs connected to a particular video input for monochrome transmissions. Color transmission imposed even more serious limitations.

Subsequent portions of this paper will describe a new relay type of video switch which extends the bandwidth considerably beyond present color requirements and removes the limitations as to the number

---

Paper 60-811, recommended by the AIEE Broadcasting Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting, Atlantic City, N. J., June 19–24, 1960. Manuscript submitted April 7, 1960; made available for printing August 15, 1960.

A. E. Bachelet, C. A. Collins, and E. R. Taylor are with the Bell Telephone Laboratories, Mr. Bachelet and Mr. Collins in Murray Hill, N. J., and Mr. Taylor in North Andover, Mass.

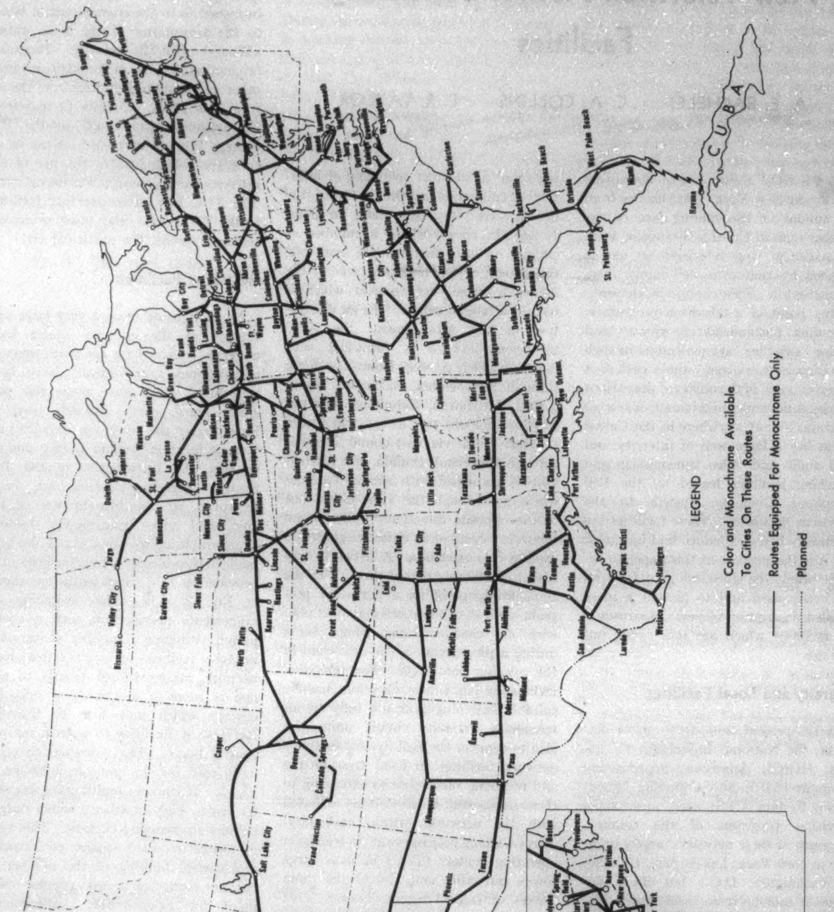

LEGEND
Color and Monochrome Available To Cities On These Routes
Routes Equipped For Monochrome Only
Planned
CUBA
Havana
Miami
West Palm Beach
Daytona Beach
Orlando
Tampa
St. Petersburg
Jacksonville
Atlanta
Birmingham
Montgomery
Mobile
New Orleans
Memphis
Nashville
Louisville
Charlotte
Richmond
Washington, D.C.
Baltimore
Philadelphia
New York
Boston
Providence
Albany
Bangor
Portland
Cleveland
Detroit
Chicago
Milwaukee
Green Bay
Minneapolis
St. Paul
Duluth
Superior
Fargo
Valley City
Bismarck
Sioux Falls
Sioux City
Omaha
Lincoln
Des Moines
Kansas City
St. Louis
Tulsa
Oklahoma City
Dallas
Fort Worth
Houston
San Antonio
Austin
Corpus Christi
Laredo
El Paso
Amarillo
Lubbock
Albuquerque
Denver
Colorado Springs
Casper
Salt Lake City
Grand Junction
Phoenix
Tucson
Yuma
Los Angeles
San Diego
Tijuana
Bakersfield
Fresno
Sacramento
San Francisco
Oakland
San Jose
Stockton
Salinas
Santa Barbara
Chico
Medford
Eugene
Portland
Seattle
Yakima
Spokane
Las Vegas

Fig. 1. Intercity Bell System television facilities

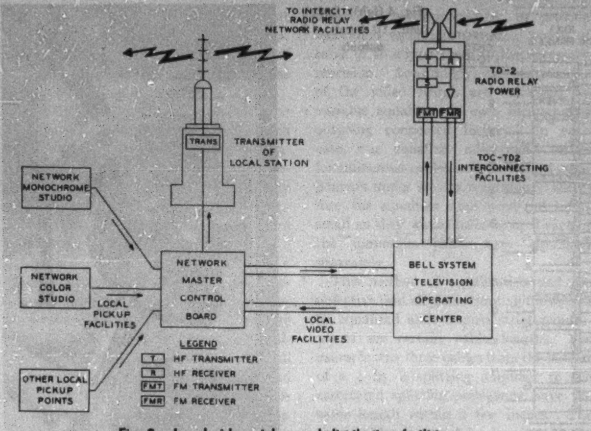

TO INTERCITY RADIO RELAY NETWORK FACILITIES
TD-2 RADIO RELAY TOWER
T
R
S
FMT
FMR
TRANS
TRANSMITTER OF LOCAL STATION
TOC-TD2 INTERCONNECTING FACILITIES
NETWORK MONOCHROME STUDIO
NETWORK COLOR STUDIO
LOCAL PICKUP FACILITIES
NETWORK MASTER CONTROL BOARD
BELL SYSTEM TELEVISION OPERATING CENTER
LOCAL VIDEO FACILITIES
OTHER LOCAL PICKUP POINTS
LEGEND
T HF TRANSMITTER
R HF RECEIVER
FMT FM TRANSMITTER
FMR FM RECEIVER

Fig. 2. Local video pickup and distribution facilities

of output circuits which may be connected to a particular program input. The integration of the new video switch into the over-all television network and the methods of controlling it will also be discussed.

NEW VIDEO SWITCH

The new video switch includes splitting pads, relay contacts, amplifiers, and cables so designed that any input may be connected to any output with negligible impairment of the video signal. Switching is on a 124-ohm balanced, constant-impedance basis with the inputs terminated at the switching relays when not in use. Fig. 4 illustrates the basic switching pattern for a video switch having 10 inputs and 12 outputs (10 by 12). Each input may be connected to any or all of the outputs by the operation of one or more relays in the switch. The method of controlling the operation of the switching relays will be discussed later. Each input is terminated in a splitting pad, shown in Fig. 5, having 12 output legs, there being about a 54-db (decibel) loss between any two output legs, to prevent a momentary open on one leg from producing an appreciable transmission effect on circuits connected to other legs of the pad. For each output circuit there is a 10 by 1 switch, containing 10 wire-spring relays. Two of these 10 by 1 switches are shown in Fig. 6, and each is capable of connecting any one of ten inputs to the output associated with the 10 by 1 switch. Each of the 12 legs of an input splitting pad is terminated in one relay in each of 12 different 10 by 1 switches.

The splitting pads and 10 by 1 switches are assembled into a basic 10 by 12 switch unit as shown in Figs. 7(A) and (B). This 10 by 12 switch unit may be completely or partially equipped as required. If less than the full complement of output switches is furnished, the unused legs on the splitting pads are terminated either at the pads or on a cable terminating panel included in the 10 by 12 switch assembly. The cabling between the splitting pads and the relays of the 10 by 1 switches is illustrated in Fig. 7(C).

Fig. 8(A) illustrates how four of the 10 by 12 switch units may be combined to form a 20 by 24 video switch, such that each input may be connected to any or all of the outputs as desired. Here each input circuit enters a splitting amplifier having three identical outputs, two of which, in the case shown, are connected to two separate 10 by 12 switch units. Each output trunk circuit includes a 3 by 1 switch, containing three wire-spring relays, two of which, in the case shown, are connected to two of the 10 by 12 switch units. With the operation of a relay in a 10 by 12 switch unit and a relay in a 3 by 1 switch unit any input may be connected to any output. An amplifier is included in each output circuit, as shown in Fig. 8(A) to make up for the loss in the input splitting pad.

Fig. 8(B) illustrates how nine 10 by 12 switch units may be combined into a 30 by 36 video switch with all inputs having access to all outputs. Here the operation of a relay in a 10 by 12 switch unit and a relay in a 3 by 1 switch serves to connect any one of 30 inputs to any one of 36 outputs.

Any number of outputs may be connected to a single input, without transmission impairment, in any of these switching patterns since the operation of a video switch simply substitutes the 124-ohm input impedance of the output amplifier for the 124-ohm impedance of the termination on the video-switch contacts. The impedance which an input "sees" looking into the video switch thus remains constant no matter how many outputs are connected. However, the switching control is so designed that no two inputs may be connected to the same output.

One or more output circuits from the new video switch are reserved for monitor positions, as shown in Fig. 9, each being capable of being switched to any input circuit. Each monitor output in turn contains a splitting amplifier making it possible to receive a program simultaneously on a monochrome and a color monitor and on a waveform oscilloscope. One or two test positions are furnished, each having an input to the video switch permitting test signals to be transmitted over any outgoing circuit.

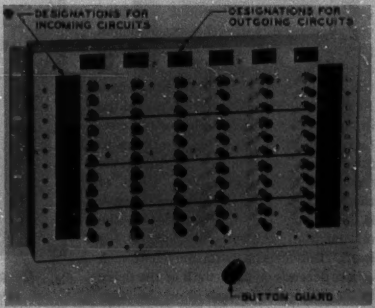

DESIGNATIONS FOR INCOMING CIRCUITS
DESIGNATIONS FOR OUTGOING CIRCUITS
BUTTON GUARD

Fig. 3. Push-button type of video switch panel

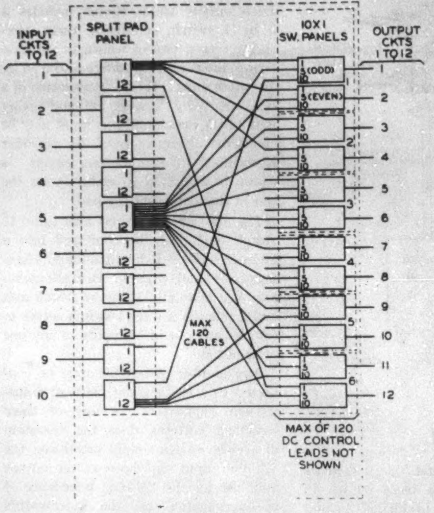


Fig. 4 (left). Switch unit, 10 inputs, 12 outputs

Fig. 5 (right). Splitting pad, 1 input, 12 outputs

Each test position also has an output circuit from the video switch permitting any incoming circuit to be connected to the receiving equipment in the test position for the reception of test signals. The improved switching flexibility and associated new testing facilities permit more rapid testing and alignment of the various components of the intercity and local video facilities.

New Video Switch Transmission Considerations

The purpose of the video switch may be stated simply as follows: to rapidly connect any incoming circuit to any one or more outgoing circuits with little or no distortion in the video signal or opportunity for error in switching. This results in each input circuit having as many branches as there are output circuits, if simple remote switching operation is desired, rather than manual plug and jack patching. For example, in a small 10 by 10 video switch each of the ten inputs would have ten branches. Each branch, including the cable from the input junction point to the switch contacts, may introduce slight echoes due to mismatches between the cable and the terminating impedance. The effect of these echoes on the wanted signal path is determined by the amount of the mismatch and the value of the leg-to-leg loss. The television distortion due to such echoes depends on the length of the delay in the echo paths as well as on the magnitude of the echo. Accordingly, in the new switch the connection from the junction point is kept electrically short and is terminated in the high-frequency image impedance of the cable. Also, the junction point and branch circuits terminate the incoming circuit in its characteristic impedance. A constant impedance splitting pad, shown in outline form in Fig. 10, is the most economical junction which meets these requirements. The attenuation requirements for the splitting pad may be stated as follows:

1. The attenuation from one output to another must be large enough to reduce the echo from the switch to negligible proportions. The requirement for this attenuation increases as 20 log N if the echoes from each branch are similar, or as 10 log N if they are random.
2. The attenuation between outputs should be at least 46 db to permit momentary removal of the termination on one leg during switch operation without noticeable change on the other outputs.
3. The attenuation from input to output should not be so great that noise in succeeding amplifiers becomes objectionable.

Noise considerations indicated that the through attenuation of the pad should not exceed about 28 db and operating features indicated the desirability of 12 outputs. Computations based on the use of a conventional H-type splitting pad illustrated in Fig. 10 indicated an echo magnitude about 55 db below the signal. This was considered to be excessive. Accordingly, a splitting pad was designed for sufficient loss between outputs where each output leg, as shown in Fig. 10,

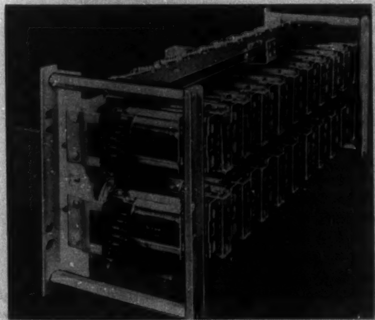


Fig. 6. Relay-type video switch panel

Vol.
Pt. I
196

contains a rectangular pad whose input shunt is the impedance of the other legs. This resulted in a further reduction of 21.5 db in echo magnitude which was considered to be adequate.

Balanced rather than coaxial circuits are used throughout the video switch because many of the TOC's are in locations where differences in ground potentials are to be expected due to nearby power, lighting, or contact making equipment; and coaxial circuits are more sensitive to these disturbances.

Each incoming or outgoing circuit in a TOC includes numerous components such as lengths of cable, splitting pads, switches, and amplifiers that require equalization. In order that one equalizer per incoming or outgoing circuit will suffice, each component is so designed that the residual distortion has a cable or "square root of frequency" shape. Typical values are as follows:

10-by-12 switch, 25 feet.
3-by-1 switch, 12 feet.
1-by-3 amplifier, 10 to 30 feet.
Output amplifier, 10 to 50 feet.

It is interesting to note that the amplifier equivalents change with vacuum-tube characteristics.

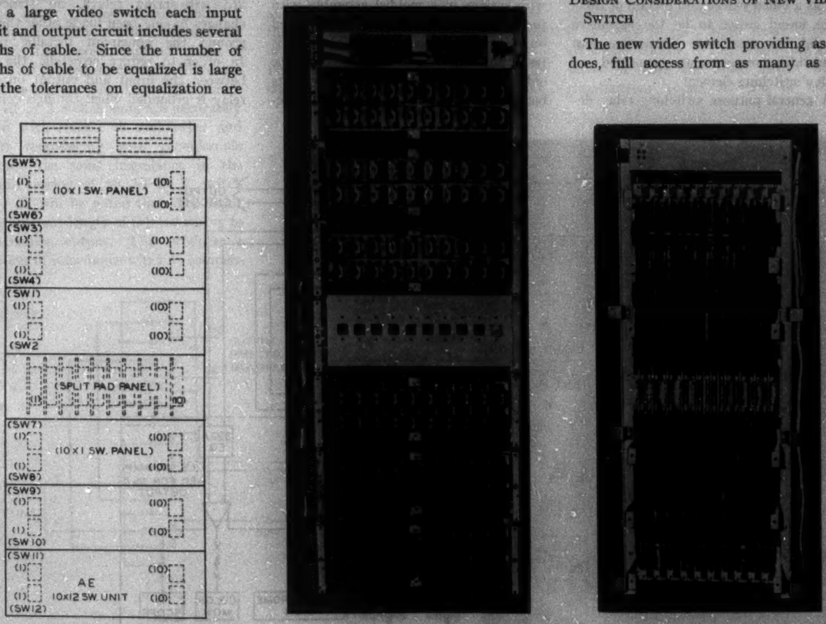
In a large video switch each input circuit and output circuit includes several lengths of cable. Since the number of lengths of cable to be equalized is large and the tolerances on equalization are severe, it is impracticable to provide individual equalizers for each length of cable or to adjust them to the necessary precision. Accordingly, the components of the video switch are equalized by variable equalizers in each incoming or outgoing connected facility. In each case, one equalizer adjustment serves for numerous cables and associated components under service conditions; therefore, the equalizer steps need not be as small as they would have to be if each of the numerous parts were equalized separately.

This method of equalization and the objective that all paths through the switch be equalized alike impose strict requirements on certain cable lengths. For example, the three cables from the output of a 1 by 3 splitting amplifier to the associated splitting pads must have the same length within a few inches. The three cables from a 3 by 1 switch to the 10 by 1 switch outputs have similar requirements.

Each incoming or outgoing facility is equalized to or from the input to a 10 by 12 switch unit, i.e., as near as feasible to the actual switch contacts. This, for discussion purposes, is referred to as point *X* and the general equalization procedure as 1-point equalization. Each incoming facility, equalized to point *X* encounters, when switched, an outgoing facility equalized from that point. In addition to the advantages discussed, this arrangement simplifies testing. Access through the video switch to connected facilities for testing purposes is obtained by means of transmitting and receiving testing facilities, which are themselves equalized to point *X*. The test equipment, therefore, functions as though it were located at the input to the 10 by 12 switch unit and terminated at the connected facility in the same impedance which a through connection encounters. Fig. 9 illustrates this arrangement. Patching facilities, using test connecting circuits having carefully cut cable lengths, are used to permit transmitting or receiving test gear to be quickly connected through the video switch to any service facility. Monitor and oscilloscope connecting circuits are also available for observing incoming signals or test patterns.

Design Considerations of New Video Switch

The new video switch providing as it does, full access from as many as 30



(A) (B) (C)

Fig. 7. Video switch unit, 10 inputs, 12 outputs (A); relay side (B); wiring side (C)

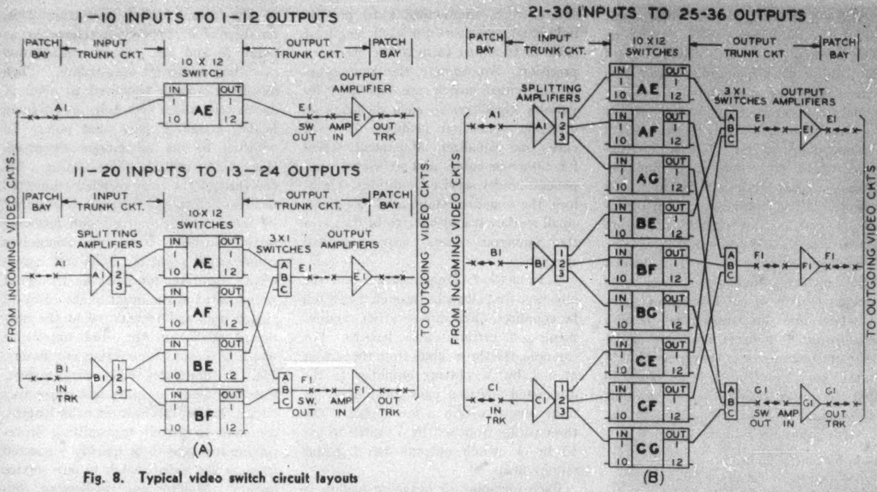

1–10 INPUTS TO 1–12 OUTPUTS
PATCH BAY
INPUT TRUNK CKT.
10 X 12 SWITCH
OUTPUT TRUNK CKT.
PATCH BAY
OUTPUT AMPLIFIER
A1
IN 1 10
AE
OUT 1 12
E1
SW OUT
AMP IN
OUT TRK
11–20 INPUTS TO 13–24 OUTPUTS
10 X 12 SWITCHES
SPLITTING AMPLIFIERS
3 X 1 SWITCHES
OUTPUT AMPLIFIERS
AF
BE
BF
B1
IN TRK
F1
SW. OUT
FROM INCOMING VIDEO CKTS.
TO OUTGOING VIDEO CKTS.
(A)
21-30 INPUTS TO 25-36 OUTPUTS
AG
BG
CE
CF
CG
C1
G1
(B)

Fig. 8. Typical video switch circuit layouts

inputs to as many as 36 outputs, requires a large number of switching elements. Several different types were studied for use as a video switch and certain of them were found either to be too expensive or did not meet the attenuation and bandwidth requirements expected of such a switching element.

A general-purpose switching relay designed to satisfy automation manufacturing processes has recently been produced by the Bell System. This relay has wire springs in a molded assembly with precious-metal double contacts welded on the ends of the springs. This relay is faster and requires less power than the older flat spring types and, because of its basic design, has closely controlled characteristics which permit it to be incorporated into the video-switch design. The relays used in the video switch are standard wire-spring relays, no structural changes having been made other than to ground the noncircuit metallic parts. In order to reduce crosstalk, the link between the input and output on each relay is grounded when through connec-

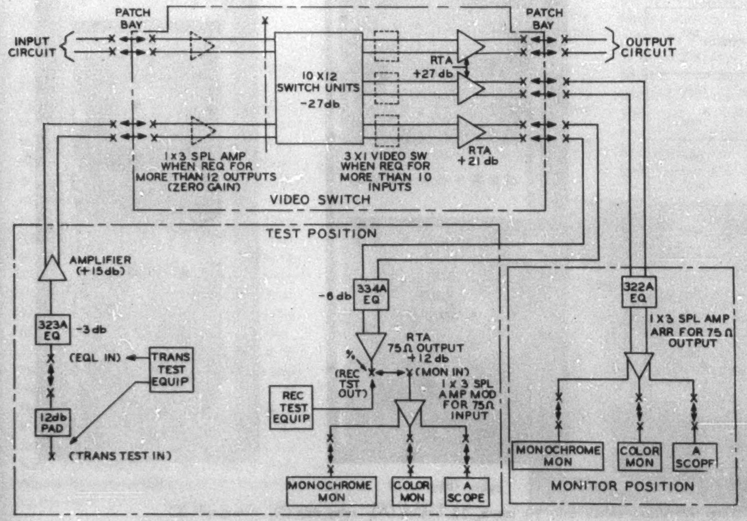

PATCH BAY
INPUT CIRCUIT
10 X 12 SWITCH UNITS -27db
RTA +27 db
PATCH BAY
OUTPUT CIRCUIT
1 X 3 SPL AMP WHEN REQ FOR MORE THAN 12 OUTPUTS (ZERO GAIN)
3 X 1 VIDEO SW WHEN REQ FOR MORE THAN 10 INPUTS
RTA +21 db
VIDEO SWITCH
TEST POSITION
AMPLIFIER (+15db)
323A EQ -3 db
(EQL IN)
TRANS TEST EQUIP
12db PAD
(TRANS TEST IN)
-6 db 334A EQ
RTA 75Ω OUTPUT +12 db
(REC TST OUT)
(MON IN)
REC TEST EQUIP
1 X 3 SPL AMP MOD FOR 75Ω INPUT
MONOCHROME MON
COLOR MON
A SCOPE
322A EQ
1 X 3 SPL AMP ARR FOR 75 Ω OUTPUT
MONOCHROME MON
COLOR MON
A SCOPE
MONITOR POSITION

Fig. 9. Test and monitor circuits

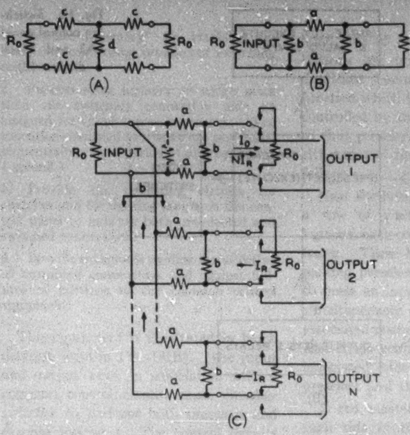


Fig. 10. Typical H-type pad (A); typical rectangular pad (B); rectangular pad arranged for switching (C)

tion is not desired. Further crosstalk reduction is obtained by using two sets of transfer springs as a ground plane between the input and the output, as shown in Fig. 11. Since all switching is on a balanced basis, two leads may be switched through each relay, one lead connecting to the upper printed circuit panel of the associated 10 by 1 switch unit and the other to the lower. Fig. 12 shows a view of the printed circuit multiple wiring. The two printed circuit multiple panels of a 10 by 1 switch unit are accurately spaced with respect to each other and permit reproducing with considerable accuracy the same impedance in the multiple wiring of all 10 by 1 switch units. It will be noted that each input has a different length of printed wiring to the common output. This produces a variation in impedance which is compensated for by series inductances of differing values. Common output coils assist in compensating for the capacitance to ground of springs on the associated wire-spring relay video switches. The series inductances make each circuit through the switch closely similar to each other circuit and provide a transmission characteristic through the switch which has a cable or square root of frequency shape. For the 10 by 12 switching unit, the transmission from any input to any output matches about 25 feet of video cable to ±0.02 db or better and up to 10 mc (megacycles).

Amplifiers

The splitting pads, described previously, introduce approximately 27-db loss into each connection between an input and an output. To make up for this loss a receiving terminal amplifier, a component of the *A2A* video systems, is introduced into each video switch output trunk circuit. This is a 3-stage video amplifier which for this application is connected so that both input and output are balanced 124 ohms. It provides about 27-db gain with a nominal bandwidth of 4.5 mc. The gain-frequency characteristic matches 10 to 50 feet of cable within 0.06 db to 4.5 mc.

The balanced amplifier stages have several advantages, the most important of which are the reduction in distortion due to balancing out the even-order distortion products of the electron tubes, and the reduction of longitudinal disturbances at the input of the amplifier.

As shown in Figs. 8(A) and (B), whenever the number of output trunk circuits exceeds 10, a 1 by 3 splitting amplifier with balanced input and output is provided in each input trunk circuit. In a video switch, a 1 by 3 splitting amplifier provides one input circuit with simultaneous access to two or three

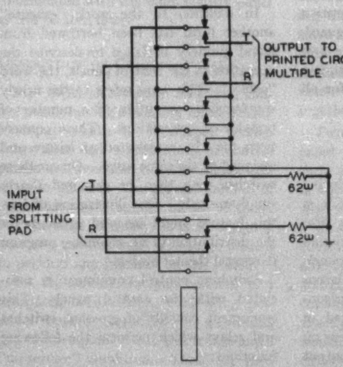


Fig. 11 (left). Method of wiring video relays of 10-input 1-output switch

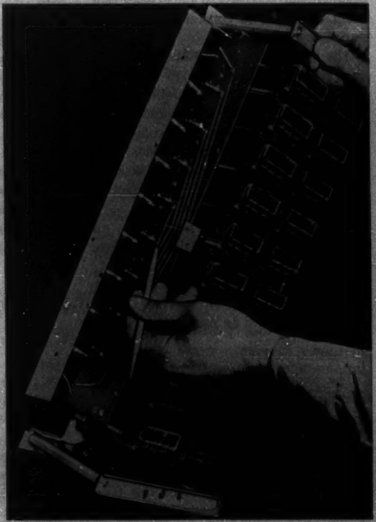


Fig. 12. (right) Video switch multiple wiring for 10 inputs, 1 output

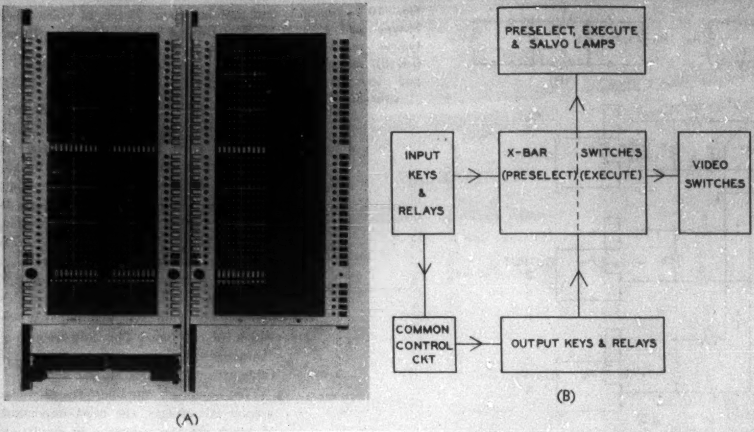


Fig. 13. Switching control panel (A) and block diagram (B)

10 by 12 switching units, each serving 12 or less outgoing trunks. The output circuit of this 2-stage balanced amplifier includes a 1 by 3 splitting pad affording 20-db loss from output to output and no net gain or loss from amplifier input to pad output. The two vacuum tubes in the output stage are so connected that the plate of one and the cathode of the other feed one side of the balanced output and vice versa. This symmetrical arrangement not only balances out even-order distortion but also effect a large saving in output power at the expense of a small reduction in constancy of output impedance.

The somewhat variable output impedance and the relatively low leg-to-leg attenuation are not objectionable for this amplifier because the three outputs are permanently terminated in accurate terminations (1 by 12 splitting pads) and the connecting cables are short (35 feet). This amplifier matches 10 to 30 feet of cable within 0.06 db to 8 mc.

The same amplifier, except with an unbalanced output, is used in the new television monitoring bay and test-monitor bay to provide simultaneous operation of a monochrome monitor, a color monitor, and an oscilloscope.

Switching Control

The older type of video switches described earlier were contacts on multi-button keys as illustrated in Fig. 3. The operation of a single button served to connect the input of the associated horizontal row of buttons to the output of the associated vertical row. Since this type of switch had no indicating lamps, the position of an operated switch was indicated by a green plastic guard ring which was placed over the operated button and flush with the top of the button. An unoperated button was protected against accidental operation by a white guard ring. The switches to be operated next were indicated by red guard rings placed over the buttons to be pushed. These are referred to as "preselected" switches. Leisurely preselection in advance of actual switching to reduce operating errors.

In addition to video switching, a TOC has control over remotely located i-f (intermediate-frequency) switches at TD radio rooms or towers and switches in coaxial terminals. Buttons to operate these switches over remote-control systems are located on monitor positions in the TOC. Each button, when operated, causes an "order" to be sent over the remote-control system for the operation of the selected switch. As a safeguard, it is usually necessary to operate two buttons for each order, one of which is a master button, used in common for all orders.

New Switching Control System

As part of the plan for introducing improved video switching in TOC's, a new switching control system is also being furnished. The video switches are operated through main control panels having a co-ordinate pattern of lamps and associated keys arranged in horizontal and vertical rows as illustrated in Fig. 13(A). The input trunks appear on the horizontal rows and the output trunks on the vertical. The intersection of a horizontal row with a vertical row represents a switching connection from the associated input to the corresponding output. A switching connection from an input to an output may be "preselected" prior to the time the switch is actually operated. A white preselect lamp and a red "execute" lamp are furnished at each co-ordinate point on the panel. The condition of each lamp of the pair describes the various states in which the possible connection between the input and the output represented may be. These states are as follows: (1) neither lamp lighted, no connection between input and output and none immediately expected; (2) white lamp lighted, connection preselected but not yet executed; (3) red lamp lighted, connection existing between the input and the output represented.

In addition to the work "execute" another term has been borrowed from naval gunnery parlance to describe the operation of the control panels, the word "salvo." This term refers to the nearly simultaneous execution of a number of preselected connections. These connections may be preselected at leisure and assigned to the same salvo. On cue these switches may then be executed nearly simultaneously. One illustration of how this feature might be used would be in the distribution of an incoming program to several theater outlets.

Switching control equipment is associated with the control panels. This equipment consists of crossbar switches and relays which perform the following functions:

1. Provide "memory" for retaining information concerning preselected switching connections.
2. Furnish salvo memory to which more than one switching connection may be assigned for simultaneous execution. (The executions assigned to a salvo are performed sequentially over a period of approximately 1 second.)
3. Provide the operating circuits for switches and for lighting lamps on the control panel to indicate both preselected and switched connections.
4. Provide safeguards against the selection of unwanted connections and against failures of portions of the common control equipment.

This equipment is illustrated in block diagram form in Fig. 13(B). The input and output keys, in association with a common control circuit, cause crossbar switches to perform both preselect and execute functions. The control circuits are furnished in duplicate to lessen the possibility of failure.

To correspond to each co-ordinate point on the control panel there are two crossbar switch "cross points," one of which furnishes preselection memory and the other an operating circuit for the associated switch and the corresponding red execute lamp. This switching control system, in addition to its use for operating video switches, may be used to operate remotely located i-f or coaxial switches over present remote-control systems. It may also be used for audio program switching control.

Monitor and test positions in the TOC are furnished with keys with which they may operate video switches to select incoming trunks for monitoring or testing on incoming circuits. Test positions have additional keys which permit selection of outgoing circuits for transmitting tests. The operation of these keys is co-ordinated with the main control panel as will be seen later.

Description of Control Panel

Fig. 13(A) shows main switching control panels having a capacity for switching 30 input trunks to 30 output trunks. Other sizes and combinations of control panels are available but this figure will serve to illustrate the general method of operation which applies to all of them. For the sake of safety, it is so designed that two buttons must always be pushed to perform any function on the panel.

On the left of the panels there is a maximum of 31 black, nonlocking preselect buttons arranged in a vertical row to correspond to 30 input trunks and a "no-service" condition. These are used in making preselections. On the right-hand side of the panel 31 similar push buttons are provided. These cause a duplicate common control circuit to function which is independent of the one controlled by the buttons on the left side so that preselections may be made from either side of the panel, in case of circuit failure.

Near the bottom of the panels there is a row of white nonlocking "output" buttons, each corresponding to an output trunk. These buttons are also used in making preselections, it being necessary to press an input and an output button simultaneously in order to preselect the associated connection. Suitable designation strips are furnished to indicate the associated video circuits for both the preselect and the output buttons.

A red master button is furnished at each side of the panel for use in the execution of preselected connections. In the lower portion of each panel are located five orange colored salvo buttons and associated amber salvo lamps, there being one amber lamp per output trunk in each salvo row. As each connection is preselected on the panel, it is automatically assigned to the top salvo and the amber lamp corresponding to the associated output is lighted in that row. A preselected connection may be assigned to another salvo row by pressing the associated output button and the button of the desired salvo simultaneously. The amber lamp corresponding to the output will go out in the top salvo row and a new lamp corresponding to the output will be lighted in the new salvo row. Any switch may be individually executed from any salvo row by pressing the master button and the associated output button simultaneously. All the switches assigned to any salvo may be executed nearly simultaneously by pressing the master button and the desired salvo button simultaneously. Two or more salvos may be executed nearly simultaneously if desired by pressing the master button and the associated buttons together.

Two green "annul" buttons are located at the bottom of the panel and these are wired in parallel for convenience in operation. If, through error or otherwise, a line preselection is made such that a white preselect lamp is lighted at the co-ordinate point where a red lamp is already lighted, meaning that this outgoing line is to be switched to the incoming line it was connected to previously, it should be noted that there would be a momentary opening of this circuit as the control circuit functions. At the time of a network break this may not be of any concern. However, if this program is a continuing one while other programs are being switched, this particular switch should not be executed. This undesired white lamp and its associated amber lamp in the salvo bank can be extinguished and the preselection nullified by holding in operation the green annul button and pushing the associated white outgoing line button.

In order to release outputs from service and make them available for testing, the lowest horizontal row of preselect and execute buttons on the main control panel is designated as a no-service row. The execute lamps in this row are green. When an output circuit is not in use, it may be preselected and executed to this row in the same manner as to any one of the 30 regular input rows. As soon as the execute or green lamp corresponding to an output circuit lights in the no-service row, an "available" lamp corresponding to the output circuit will light in each test position. The output circuit may now be selected by either of the test positions but control of it remains at the main control panel so that it may be reseized for program service at any time. The operation of an output selection key on the test position also lights a "test in progress" lamp corresponding to the output circuit on the main control panel. These lamps have novelty caps and are located on the main control panel in a horizontal row just below the no-service lamps. One of these lamps lights at any time the associated key in any test position is operated so that, in addition to indicating an output circuit under test, it may be used to indicate that the test position wishes to obtain an output circuit not yet released for test.

Remote Switching Control Equipment

As explained earlier, there are many instances where a TOC in order to establish a particular connection, is required to operate one or more i-f switches in a remote radio tower or coaxial terminal as well as video switches in the TOC. Where this type of operation is required, one portion of the control panel can be assigned to the operation of the i-f or coaxial switches over a remote-control system and another portion to the local video switches. Either switching system can be operated independently on an "individual execute" basis or switches of both systems may be operated simultaneously as part of the same salvo.

The new switching control panel has been arranged to work either with existing remote switching control systems or with a new remote-control switching system designed especially to function with the new control panel. The existing systems operate the remotely located switches either over direct-current circuits or by means of sequential signaling over a voice-frequency circuit such as a call wire. No indication is given to the control center as to whether the correct switch has been operated. The new remote-control system not only orders the operation of a remotely located switch, but checks to see that the proper switch has been selected and acknowledges its operation by the lighting of associated lights at the control TOC.

The new remote-control arrangement uses a pulse-length code frequency shift system employing *43A1* carrier telegraph terminals at the transmitting, or control, end of the system and at each receiving point. The incoming and outgoing circuits at each remote point are represented by horizontals and verticals on the main control panel at the TOC and are preselected by the operation of the proper buttons, in the manner described earlier The operation of these buttons results in sending pulse-length code orders to each remote point. Only one remote station accepts a particular set of preselection orders, however, depending on the code sent. At this station the preselection is made and checked, and if the received order has been correctly handled, an acknowledgment signal is sent back to the TOC. When this acknowledgment order has been received, the lamps at the coordinate point, represented by the operated input and output buttons, light. This may be as long as 2 seconds after the buttons are pushed initially.

At the remote point, crossbar switch equipment is provided, which, under the control of the orders from the remote system, furnishes the operating paths for the remote i-f switches. A control panel at each remote point permits local control in case of emergency. Normally, this panel is not in use and its lamps do not light as remote orders are received. When switched to local control, however, the lamps on the remote panel indicate preselections and executions in the usual manner. Any order sent from the control panel at the TOC will automatically restore the local panel to remote operation.

Conclusions

The new video switching equipment together with the associated new monitoring and testing facilities are part of an over-all integrated plan for furnishing broad-band transmission facilities. These may be improved, in accordance with the requirements of future developments in the color and theater television fields, up to bandwidths of 0–10 mc. The new switching control equipment offers safer and more flexible control of switching facilities, and it is expected that these will be widely used for television network services furnished by the Bell System.

Radio-Frequency Transmission Line Engineering

R. F. LEWIS
NONMEMBER AIEE

IN THE development of a transmission line, the engineer attempts to establish a criterion for minimum attenuation; minimum discontinuities, or minimum VSWR (voltage standing wave ratio); maximum power handling capability; maximum frequency range; ease of manufacture; and the ability to attach connectors having consistent and reliable electrical and mechanical qualities. This paper will discuss the development of a semiflexible, semisolid dielectric coaxial transmission line and its associated connectors, with particular reference to obtaining minimum discontinuities. A cycling method of measurement is used to ensure complete frequency coverage as well as to distinguish discontinuities between connectors and those inherent in the coaxial transmission line. The first part of the paper will consider the general aspects of the coaxial transmission line and its associated connectors; the second part will discuss a specific consideration of the problem of analyzing cycling traces in determining over-all discontinuity effects.

General Considerations

The general axiom accepted in the development of semiflexible semisolid dielectric transmission line is that the metals have excellent conductivity qualities and that they be sufficiently malleable to allow bending to a given minimum radius. The dielectric also must have good mechanical and electrical qualities; mechanically, it must support the inner conductor in a concentric position in any configuration, at least to the minimum bending radius. The problem is to choose sufficient dielectric material to keep the inner conductor concentric with the outer for all permissible configurations, yet not to the extent of causing excessive attenuation due to the dielectric. A desirable feature is to have all sizes of cable for a given characteristic impedance designed in the same manner. This feature adds to the ease of manufacture, and in particular it eases the problem of preparing the transmission line for low VSWR connectors.

In designing transmission lines for maximum frequency response, experience has shown that minimum discontinuities can be obtained from the solid-dielectric or continuous-dielectric contact type of transmission lines at the expense of high attenuation; and that minimum attenuation is obtained in discrete insulator-supported air dielectric transmission line at the expense of large discontinuities over relatively narrow frequency bands. The characteristics of semisolid transmission line places it between the solid and air dielectric types.

Spir-O-Line is a semiflexible semisolid dielectric transmission line developed and marketed by the company with which the author is associated. It is composed of a copper inner conductor and aluminum outer conductor. Six high-density polyethylene tubes are symmetrically placed about the inner conductor to make it concentric with the outer conductor. Sufficient pressure results from the sinking-down process of the outer conductor to ensure constant and consistent contact of the dielectric tubes on both inner and outer conductors. Fig. 1 shows a cross section of the 1/2-, 7/8-, $1^5/_8$-, and $3^1/_8$-

---

Paper 60-899, recommended by the AIEE Radio Communication Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting, Atlantic City, N. J., June 19–24, 1960. Manuscript submitted April 25, 1960; made available for printing May 3, 1960.

R. F. Lewis is with Prodelin, Inc., Kearny, N. J.

Vol.
Pt. 1
1960

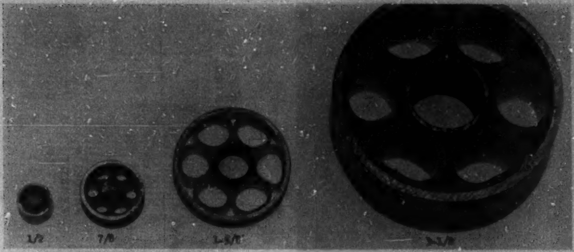

1/2
7/8
1-5/8
3-1/8

Fig. 1. Cross section of (left to right) 1/2-, 7/8-, $1^5/_8$-, and $3^1/_8$-inch 50-ohm Spir-O-Line

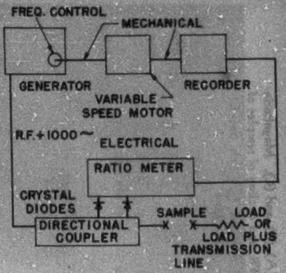

FREQ. CONTROL
MECHANICAL
GENERATOR
VARIABLE SPEED MOTOR
RECORDER
R.F. + 1000 ~
ELECTRICAL
RATIO METER
CRYSTAL DIODES
DIRECTIONAL COUPLER
SAMPLE
LOAD OR LOAD PLUS TRANSMISSION LINE

Fig. 4. Schematic diagram of equipment used in recording VSWR cycling of Spir-O-Line transmission line and connector assemblies

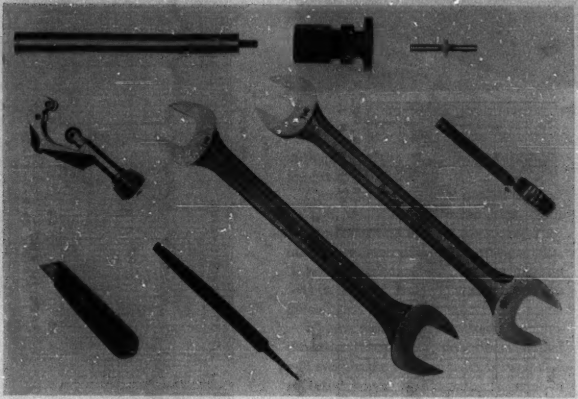


Fig. 2. Tools for trimming transmission line

inch 50-ohm Spir-O-Line. It is felt that this construction is a very close approach to a continuous dielectric wherein all portions of the transmission line have the same velocity of propagation. This type of construction permits all sizes to be made in the same manner; thus all sizes have the same capacity per unit length, and therefore the same phase velocity. This feature aids in the ability to design connectors since the dielectric interface effect will be the same for all cable sizes.

A further advantage of this type of construction is that it permits Teflon tubes to be used as the dielectric, whereby temperature rating is increased approximately two and one half times.

The transmission line should be so designed that standard hardware tools can be employed in preparing the ends for various connectors. Certainly, if such is the case only the knowledge of such tools is required to dress the line for connectors. Further, it is desirable to be able to cut the dielectric perpendicular to the axis of the conductors. The tools for trimming the transmission line are shown in Fig. 2, where it may be seen that the only tools necessary are tube cutter, razor knife, measuring tape, file, and wrenches. The feature of cutting the dielectric perpendicular to the axis is particularly desirable in eliminating discontinuities in the high-frequency spectrum, or in the event that it is desired to have many lengths where each length shall have the same electrical characteristics.

All connectors developed for use with Spir-O-Line have the same basic grip design where they couple to the sheath, and are known commercially as Spir-O-Lok connectors. A typical cutaway of a Spir-O-Lok connector is shown in Fig. 3.

Two mating tapered sleeves achieve a positive grip. The holding sleeve has grooves machined at an angle designed to shear into the cable sheath to a predetermined depth. This depth multi-

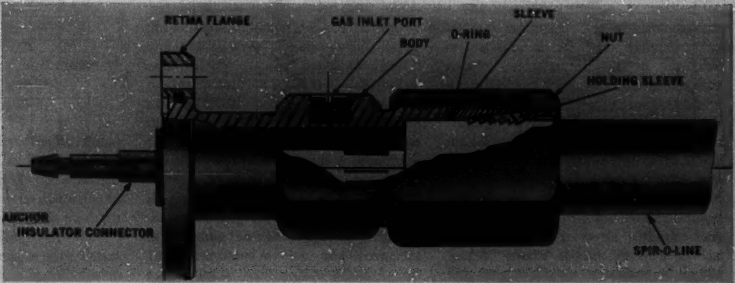
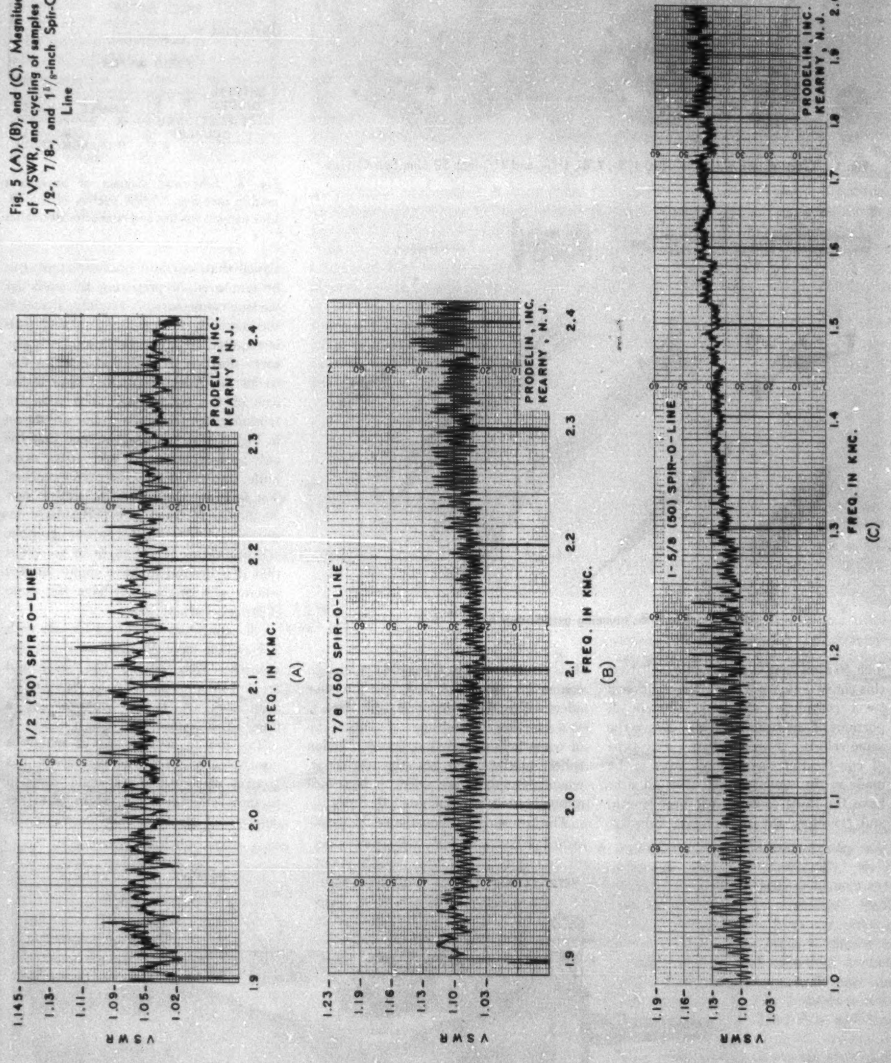

RETMA FLANGE
GAS INLET PORT
BODY
O-RING
SLEEVE
NUT
HOLDING SLEEVE
ANCHOR
INSULATOR CONNECTOR
SPIR-O-LINE

Fig 3. Typical cutaway of Spir-O-Lok connector

Fig. 5 (A), (B), and (C). Magnitude of VSWR, and cycling of samples of 1/2-, 7/8-, and $1^5/_8$-inch Spir-O-Line



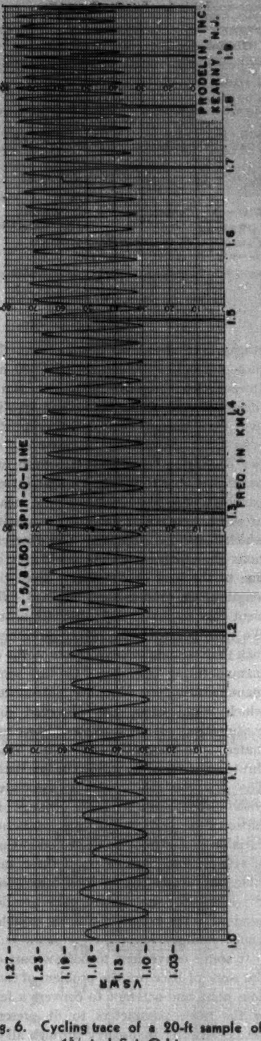


Fig. 6. Cycling trace of a 20-ft sample of $1^5/_8$-inch Spir-O-Line

plied by the number of grooves will equal the thickness of the sheath; thus a Spir-O-Line connector when properly attached to the outer sheath causes no distortion to it, and the connector has a tensile strength equal to the sheath itself.

All Spir-O-Line connectors have Teflon insulation, thereby eliminating confusion in their use with low-temperature or high-temperature cables.

Fig. 7. Test equipment

There is always a great deal of interest in how the different parameters of a transmission line and its associated connectors are determined at radio frequencies. A brief description of a cycling method over a limited frequency range will be given.

When a radio wave is passed through a directional coupler, the directional coupler will sample the forward wave as well as a reflected wave, if present. If, now, a given length of transmission line is connected to the directional coupler where the far end is terminated in the characteristic impedance of the transmission line, then, under ideal conditions there should be no reflected component. If, however, there are discontinuities or mismatches either at the terminal, or within the transmission line there will be a reflected component in the directional coupler. If the radio frequency is allowed to be continuously swept over a given range, then discontinuities due to the terminal end will produce a periodic response. It is evident that many combinations of discontinuities and transmission line lengths can be arranged to give periodic cycling. For example, with matched connectors, and a good terminal load equal to the characteristic impedance of the transmission line, there should be evidence of the cycling rate due only to small discontinuities at the input and terminal ends with perhaps some evidence of random cycling due to the transmission line. Fig. 4 is a schematic diagram of the equipment used in recording the VSWR cycling of Spir-O-Line transmission line and connector assemblies.

Figs. 5(A), (B), and (C) indicate the magnitude of the VSWR, and the cycling rate of representative samples of 1/2-, 7/8-, and $1^5/_8$-inch Spir-O-Line. Since the velocity of propagation of the different diameter lines is equal, the cycling rates of equal lengths are the same. It is important to observe that the magnitude of the VSWR is fairly constant, and the cycling is due to the ends of the transmission line or the end connector and load. This type of cycling indicates that the transmission line is free from internal discontinuities.

Accurate measurements of the cycling rate are useful in determining the electrical length of a transmission line, and also its effective dielectric constant. Fig. 6 shows a cycling trace of a 20-ft (foot) sample of $1^5/_8$-inch Spir-O-Line. Note that the cycling rate is now well defined, but the peaks of the VSWR are higher than those shown in Fig. 5(C). This is due to reduced attenuation in the shorter length, thereby allowing a greater amount of the reflected component at the terminal to appear at the directional coupler.

This suggests a convenient method for testing connectors. Let us select a length of transmission line sufficiently short to reduce the attenuation to a negligible quantity, but a length to produce a convenient cycling rate. Suppose that a long length of transmission line is used which masks the terminal discontinuities as a load for a short length of transmission line of the same size. If the same input and output connections are maintained, and the short length with the sample connectors are placed in series with the long length, effects should be seen of the long-length cycling superimposed on the

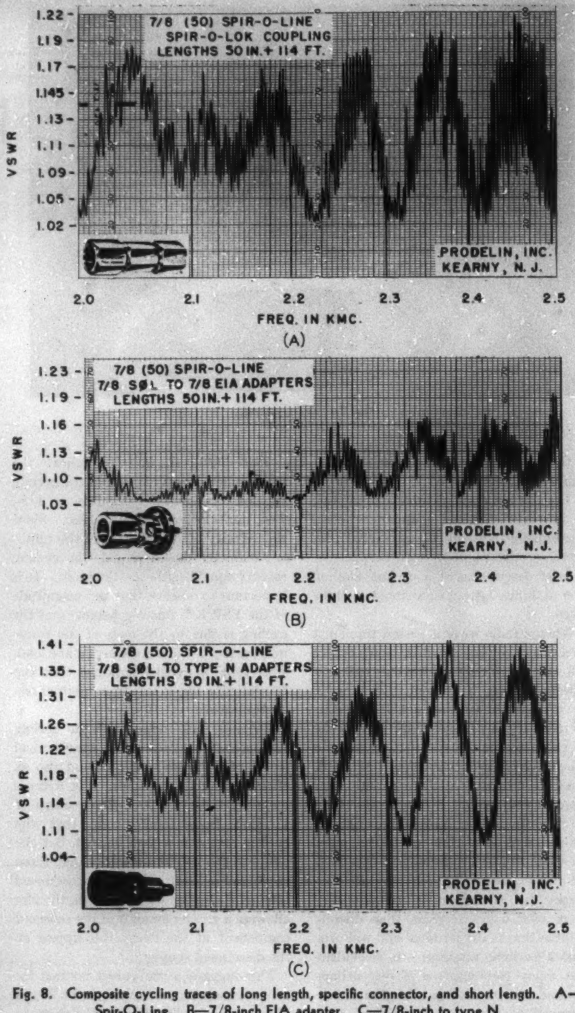


Fig. 8. Composite cycling traces of long length, specific connector, and short length. A—Spir-O-Line. B—7/8-inch EIA adapter. C—7/8-inch to type N

short-length cycling. The test arrangement for the long length, sample connectors, and short length is shown in Fig. 7. Figs. 8(A), (B), and (C) show the result of tests performed on 7/8-inch cables for a 7/8-inch Spir-O-Line coupling, two 7/8-inch Spir-O-Line to 7/8-inch *EIA*, and two 7/8-inch Spir-O-Line to type *N* connectors respectively.

In the foregoing is presented a method with rapid continuous frequency coverage for observing discontinuities in either the coaxial transmission line, the connectors, or both.

Specific Considerations

The development of Spir-O-Line presented the usual headaches. A description of one of these difficulties may be of interest as to how the condition arose, and what was done to remedy it. In the early stages of Spir-O-Line development the measurements extended up to 1,000 mc (megacycles). Above 1,000-mc measurements seemed to be somewhat erratic, and were passed over as experimental errors in measurement. However, as our measuring techniques improved, we were soon able to determine that in one particular size some of the erratic measurements were occurring with a regularity which demanded attention. Our observations showed that cycling could not be due to the load or connectors since the length of the transmission line had sufficient loss to mask their discontinuities. We were left with the rather disturbing fact that we might have the problem of manufacturing tolerance. Or had some manufacturing procedure been introduced which could cause the observed discontinuities? If the load should be eliminated, the discontinuities must exist due to some variables within the line itself. The first variable generally associated with discontinuities is the change in capacity in portions of line. Another variable that cannot be neglected is that of attenuation. Let us see if these variables are dependent on each other, and if possible to what extent they are dependent to the extent of causing troublesome cycling.

Table I. Transfer Parameters

Transmission Line	Noise
x = meter	t = time, sec
$C(x)$ = farad	$I(t)$ = amplitude
S = cycles per meter, "sinuosity"	f = frequency in cycles/sec
$w(s)$ = power spectrum	$w(f)$ = power spectrum
Mean square deviation, farads²	mean square deviation, amperes²
$(\overline{C}(x))^2=\int_0^\infty w(s)ds$	$(\overline{I}(t))^2=\int_0^\infty w(f)df$
$C(x)=\sum_{n=1}^{\infty} C_n \cos(2\pi S_n x+\phi_n)$	$I(t)=\sum_{n=1}^{\infty} C_n \cos(2\pi f_n t+\phi_n)$
$\overline{C}_n^2=2w(S_n)\Delta s$	$\overline{C}_n^2=2w(f_n)\Delta f$
$S_n=n\Delta s$	$f_n=n\Delta f$

It seems convenient to borrow part of the work of Rice[1] on the problem of random noise and attempt to convert a few of the noise parameters to fit present needs. Rather than carry through all of the derivations, we will simply state the change in parameters, and show the end result. Since the work of Rice was carried through from a power spectrum basis, the same will be done here; see Table I, a table of transfer parameters. When these transfer parameters are inserted, the following equation results:

$$|(I_r)^2|=\frac{\pi}{16}\frac{1}{\lambda_0}\left(\frac{f_0}{f_b-f_a}\right)\left(\frac{C(x)}{C}\right)^2\frac{1}{\alpha}$$

Vol.
Pt. 1
1960

where

I_r = reflected current due to discontinuities expressed as 100% of incident current

λ_0 = wavelength at center band of discontinuities, meters

f_0 = frequency at center band of discontinuities, mc/sec (mc per second)

$f_b - f_a$ = frequency range of discontinuities, mc/sec

$C(x)$ = capacity discontinuity, farads/meter

C = capacity between inner and outer conductor under ideal conditions, farads/meter

α = attenuation, nepers/meter

In this equation such terms as $(1-\epsilon^{-4\alpha l})$ were allowed to approach unity, which is permissible when long lengths in the microwave spectrum are considered.

After reviewing some of our observations, the question will be considered as to how to use the equation to determine such quantities as $[C(x)/C]^2$ and I_r. Fig. 9(A) shows a 20-ft section cut from sample *A* and cycled over a range from 2.0 to 2.5 kmc (kilomegacycles). Note a slight peak at 2.05 kmc having a VSWR of 1.19, and another slight peak at 2.35 kmc of some 1.13 VSWR. Consider the first peak. The following quantities are known:

f_0 = 2,050 mc/sec
λ_0 = 0.146 meters
$f_b - f_a$ = 25 mc/sec
$C = 77.3 \times 10^{-12}$ *fd/m*, a measured value
$1/\alpha = 1.28 \times 10^3$, a measured value taken on 100-ft sample
VSWR = 1.19
I_r = 0.0825

Inserting these values in the equation and determining the ratio $[C(x)/C]^2$ yields

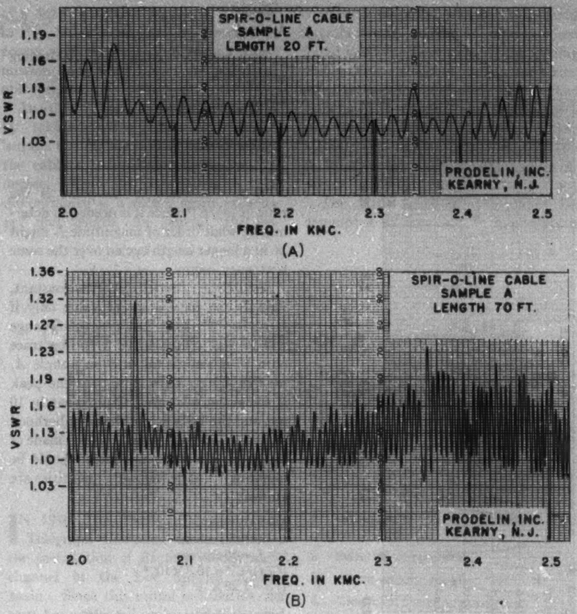
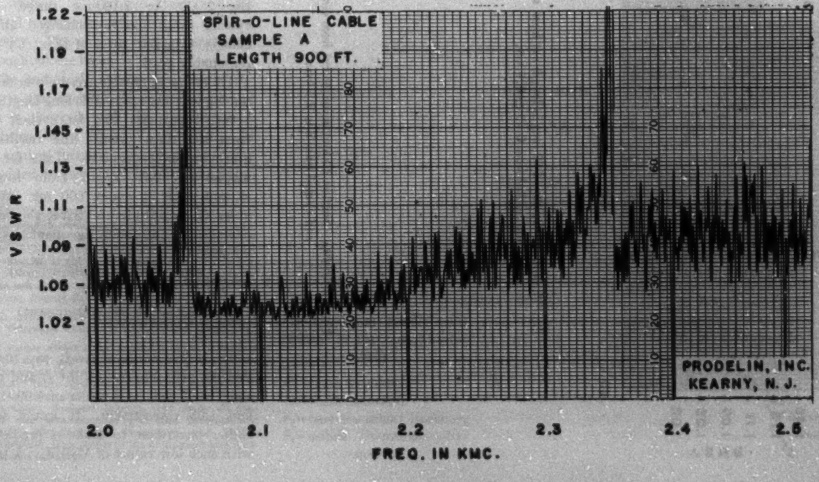
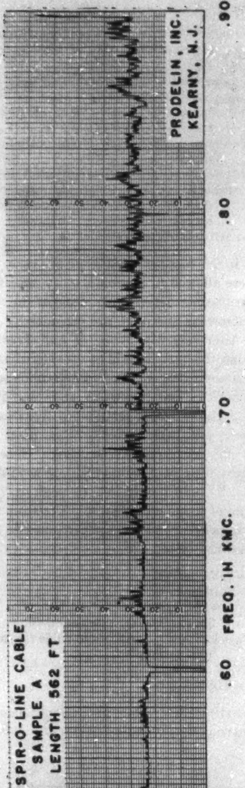


Fig. 9. Sections cut from sample A and cycled from 2.0 to 2.5 kmc. A—20-ft section. B—70-ft section

Fig. 10 (below). Section 900 feet in length cut from sample A and cycled from 2.0 to 2.5 kmc





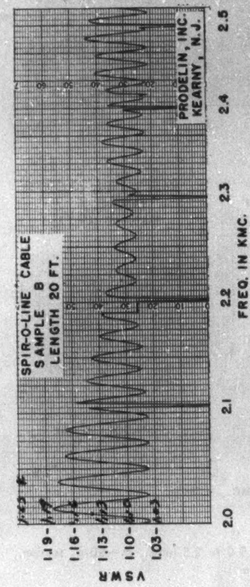


Fig. 11 (above). Section 562 feet in length cut from sample A and cycled from 0.45 to 0.9 kmc

Fig. 12 (right). Section 20 feet in length cut from sample B and cycled from 2 to 2.5 kmc

$$\left(\frac{C(x)}{C}\right)^2=16.3\times10^{-8}$$

Reducing the equation and substituting for C yields

$$\frac{C(x)}{C}=4.02\times10^{-4}$$

$$C(x)=4.02\times10^{-4}\times77.3\times10^{-12}$$

$$C(x)=0.031\times10^{-12}\ fd/m$$

Of more interest at this point is the ratio $[C(x)/C]^2$, since it is needed in determining what order of magnitude I_r might be in a longer length cycled over the same frequency range.

Should the ratio $[C(x)/C]^2$ be constant, then I_r can have a larger value only if f_b-f_a becomes less; i.e., the cycling range will be less. To observe a cycling trace on a 70-ft sample taken from sample A, see Fig. 9(B). Here, note that the peak is somewhat narrower, approximately 10 mc, and the VSWR is in the neighborhood of 1.32. If a good guess has been made of the ratio $[C(x)/C]^2$, agreement should be found in I_r. The following quantities are known:

$f_0=2{,}050$ mc/sec
$\lambda_0=0.146$ meters
$f_b-f_a=10$ mc/sec
$[C(x)/C]^2=16.3\times10^{-8}$
$1/\alpha=1.28\times10^{-3}$

Putting these values in the equation yields $I_r=0.131$, or a VSWR of 1.32. This seems too good to be true!

If lengths much greater than 70 ft are cycled, and the cycling range of the peak remains close to 10 mc, can we feel reasonably sure that we have made a good guess of the ratio $[C(x)/C]^2$? Fig. 10 shows a trace of 900 feet of sample A including the 2,000- to 2,100-mc range. The cycling range has not changed appreciably; thus our guess of $[C(x)/C]^2$ was reasonably good.

This same approach may help in determining whether reasonable long lengths at lower frequencies can be cycled, and from the trace predict the results at much higher frequencies. Take for example a 562-ft length from sample A and cycle it about 500 mc. What order of magnitude of VSWR should be observed? The following quantities are known:

$f_0=500$ mc/sec
$\lambda_0=0.6$ meters
$1/\alpha=2.64\times10^3$ (from measurement)
$[C(x)/C]^2=16.3\times10^{-8}$

Try a few values of f_b-f_a, say 10, 20, and 40 mc. The values for I_r and thus the VSWR will be in order of 1.09, 1.07, and 1.05 respectively. It would seem rather hazardous to make a prediction with such low values of VSWR. A trace of a 562-ft length of sample *A* is shown in Fig. 11, and one may be tempted to accept such a trace as indicating a transmission line as not possessing objectionable peaks.

An attempt has been made to show that by taking cycling traces on various lengths the capacity discontinuity within the cable can be determined to a fair degree. It has also been shown that the discontinuity is a rather small capacity per unit length. In general, the discontinuity discussed was due to a small manufacturing tolerance which, when found and remedied, reduced the cycling within the transmission line to a minimum. After proper manufacturing changes had been made, cycling traces were taken on a new production sample, *B*. Figs. 12 and 5(B) show such traces on a 20-ft and a 114-ft length. No cycling peaks were observed in either section.

The latter part of the paper has attempted to show how it is possible to overlook troublesome discontinuities in samples of short lengths at high frequencies, or in long sample lengths at lower frequencies. The theory was injected to show order of magnitudes which might be expected for the various dependent variables.

Reference

1. Mathematical Analysis of Random Noise, S. O. Rice. *Bell System Technical Journal*, New York, N. Y., vol. 23, 1944, pp. 282–332; vol. 24, 1945, pp. 46–156.

Centralized Testing of a Very-High-Frequency Mobile Radio System

D. L. MacDONALD
ASSOCIATE MEMBER AIEE

In 1946 The Pacific Telephone and Telegraph Company made plans for the installation of its first mobile radio channel in the Los Angeles (Calif.) basin. Since this initial installation the area has evidenced an insatiable demand for this type of service. By the end of 1960 the telephone companies serving the area plan to have all available channels in service. Fig. 1 is a block schematic of a typical channel in the Los Angeles area.

In the company's Los Angeles System, all base transmitters are located either at Mt. Wilson or Verdugo Peak and all control terminals are located in Los Angeles. The receiver sites are scattered throughout the basin.

The technical operator for these channels is stationed at the control terminal and is charged with the usual responsibilities of a licensed operator. In the control terminal equipment has been provided to perform the following functions:

1. Off-on switching of the remote transmitter.
2. Selection of receivers.
3. Alarm circuits for the transmitter.
4. Control of the test transmitter.
5. Audio level control for transmitter and receiver.

In the initial design of these channels, a frequency monitor was provided which continuously monitored the transmitter. This device gave an alarm and connected the transmitter to a dummy load whenever the transmitter drifted too far off frequency. One major trouble developed immediately. The monitors were less stable than the transmitters. As the stability of transmitters improved rapidly the need for monitors apparently disappeared and they were no longer installed. In retrospect, it is evident that some means of monitoring other than aural should have been retained.

The problem which now arises is that of a radio spectrum crowded beyond anything believable 10 years ago. Errors in operation which could be accepted then are intolerable today. Portable frequency meters now have a precision that 10 years ago could only be achieved in laboratory instruments, and still they are not good enough.

In Los Angeles this problem is further complicated by the topography of the area. Mt. Wilson has a height above average terrain of 2,800 feet as defined

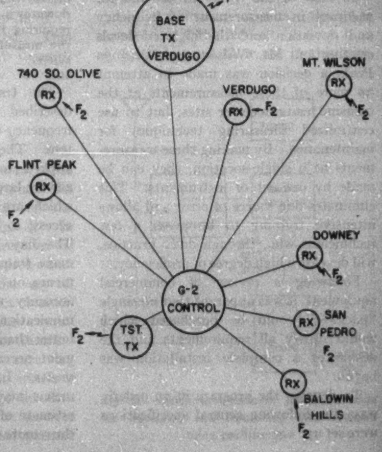


Fig. 1. Typical mobile radio channel in the Los Angeles basin

RX = base receiver
TX = base transmitter
F_1 = base transmitter frequency
F_2 = mobile transmitter frequency

Paper 60-996, recommended by the AIEE Radio Communication Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Pacific General Meeting, San Diego, Calif., August 8–12, 1960. Manuscript submitted April 1, 1960; made available for printing August 2, 1960.

D. L. MacDonald is with The Pacific Telephone and Telegraph Company, Los Angeles, Calif.

The author wishes to acknowledge the co-operation of his many associates who participated in the development of the test bay and methods, in particular: Gene Applegarth, who prepared many of the sketches and the final transmission recommendation; Emmett Ingram, who made all contacts and arrangements with the International Crystal Manufacturing Company; and Phillip Rheinschild, who prepared the equipment specifications and drawings.

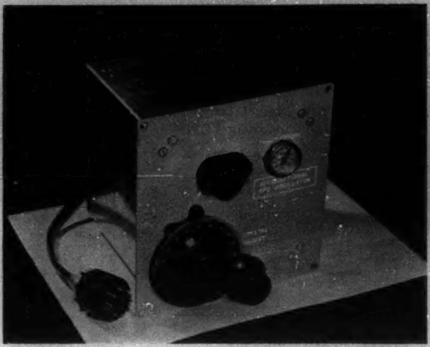


Fig. 2. Crystal turret used in multichannel receiver

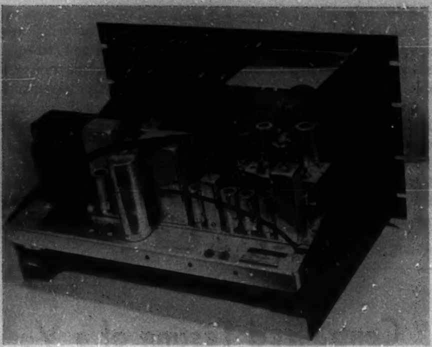


Fig. 3. Crystal turret mounted in receiver

by Federal Communications Commission rules.[1] From this standpoint it is an ideal site for the location of radio transmitting and receiving equipment. However, the same natural laws which govern site selection for mobile radio also apply to site selection for frequency-modulation and television entertainment broadcast purposes. Consequently, most major frequency-modulation and all television stations have located their transmitters on Mt. Wilson. The Pacific Telephone and Telegraph Company therefore finds it necessary to operate and maintain mobile radio equipment consisting of transmitters with comparatively modest power output and receivers of high sensitivity in a radio-frequency (r-f) field of high intensity.

In view of the increased demand for accuracy in measurement of frequency and deviation and the high r-f levels existing on Mt. Wilson and Verdugo Peak, a decision was made to attempt no more of these measurements at the transmitter or receiver sites, but to use centralized measuring techniques for maintenance. By making these measurements at a single location, they can be made by one set of instruments. This eliminates one source of error and allows intensive training to be given a few technicians who, through daily practice, will develop a high degree of proficiency.

Following a review of commercial equipment, it was apparent that no single "package" could be purchased which would satisfy all requirements and the design of a complete installation was begun.

To develop the program in an orderly way, the following general specifications were set up:

1. The design of the installation must be such that it could be readily changed.
2. A means of calibrating all instruments must be included.
3. Making the measurements must require a minimum of time.
4. Frequency measurements are required in the 35-, 43-, 152-, 157-, 454-, and 459-mc (megacycle) range.
5. Measurement must be made on transmitters in the same room as well as those 25 to 30 miles away.
6. Measurement of the frequency modulation (deviation) of the transmitters is required.
7. Initially, only base station equipment will be measured; eventually, mobile units will also be measured from this point.
8. Since a consistent, usable signal from stations *WWV* or *WWVH* cannot be received during normal working hours in downtown Los Angeles, any instrument requiring reception of *WWV* at the time the measurement is made cannot be considered.

The test bay and methods to be described are suitable for use in any frequency-modulation mobile radio system. They are economical only when applied to systems having a comparatively large number of transmitters and which are closely integrated or where access to the transmitter is difficult. The bay itself contains equipment to make frequency and deviation measurements on remote transmitters with an accuracy which exceeds Federal Communications Commission requirements by better than 100%. It contains all equipment necessary to calibrate the instruments. In addition, aural and visual means may be employed to obtain an estimate of the quality of the output of the remote transmitters.

Test Bay Equipment

The heart of the system consists of three communication receivers modified to permit reception of as many as 12 crystal-controlled channels on any one receiver. Thus, by using only three receivers 36 crystal-controlled channels are available.

The device that makes this possible is a 12-position crystal turret designed and manufactured by International Crystal Manufacturing Company. The specifications for the turret as drawn up by The Pacific Telephone and Telegraph Company impose the following requirements:

1. There must be provision for 12 crystal-controlled channels.
2. A minimum modification of the receiver should be necessary.
3. It should be as nearly a plug-in unit as possible.
4. A front panel control should be provided to shift the frequency of the selected crystal channel.
5. The stability and sensitivity of the receiver using this unit should approximate that of the standard communication receiver.

A second modification of these receivers consists of the addition of a 390-kc oscillator and mixer. This is used in making measurements of the frequency modulation (deviation) of the remote transmitters. It is also manufactured by International Crystal Manufacturing Company. Its use is discussed in the paragraphs on deviation measurement.

The first of the crystal turrets to be delivered, shown in Fig. 2, is for use in a 450-mc General Electric Company receiver. A padder control visible on the front panel allows the discriminator of

Vol.
Pt. 1
1960

Table I. Conversion of Frequency to Dial Setting for Gertsch FM-6

JL Channel Frequency Measurements: ±0.0005% = ±0.8 Kc

Difference Frequency, Cycles	Base Transmitter		Base Receiver	
	Frequencies, Kc	Gertsch Dial,* Mc	Frequencies, Kc	Gertsch Dial,* Mc
800 −	152509.2	38.12730	157769.2	39.44230
700 −	152509.3	38.12733	157769.3	39.44233
600 −	152509.4	38.12735	157769.4	39.44235
500 −	152509.5	38.12738	157769.5	39.44238
400 −	152509.6	38.12740	157769.6	39.44240
300 −	152509.7	38.12743	157769.7	39.44243
200 −	152509.8	38.12745	157769.8	39.44245
100 −	152509.9	38.12748	157769.9	39.44248
000	152510.0	38.12750	157770.0	39.44250
100 +	152510.1	38.12753	157770.1	39.44253
200 +	152510.2	38.12755	157770.2	39.44255
300 +	152510.3	38.12758	157770.3	39.44258
400 +	152510.4	38.12760	157770.4	39.44260
500 +	152510.5	38.12763	157770.5	39.44263
600 +	152510.6	38.12765	157770.6	39.44265
700 +	152510.7	38.12768	157770.7	39.44268
800 +	152510.8	38.12770	157770.8	39.44270

* Gertsch dial setting = frequency/4.

the receiver to be "zeroed" on a carrier that is displaced ±1.5 kc from the center crystal frequency. While this range is sufficient to allow the device to be used, its usefulness would be much greater if the range were extended.

Fig. 3 shows the turret installed in a receiver. Also visible are two meters mounted on the front panel.

The first of these meters, a conventional 0–500 microammeter, is connected to the first limiter test jack and to the last intermediate-frequency (i-f) output. With this arrangement it is possible to obtain a relative indication of the output of the remote transmitter. This is not an absolute indication of power out, but by noting the first limiter reading when the transmitter is putting out its rated power and the reading when the power is reduced 3 decibels, a useful indication of the power output is available, provided the sensitivity of the receiver remains constant. A means of checking this is provided in the test bay.

The second of the two meters, a 50–0–50 microammeter connected to the discriminator test jack, is used when making frequency checks.

Frequency measurements of the remote transmitter are made using one of the receivers described and an *FM-6* Gertsch frequency meter. The *FM-6* has been modified by substituting a new type of crystal oven which is expected to substantially improve the already excellent performance of the instrument. The primary advantage of the new oven is the elimination of the off-on effect of the oven thermostat. The new oven has a proportional control.

The third major item of equipment included in the test bay is a standard oscilloscope. The only major requirement is that the oscilloscope have a d-c input. There is no requirement for wide-band operation. The instrument selected for this installation is manufactured by the Hewlett-Packard Corporation and is used with no modification.

A *WWV* receiver is provided in the bay for calibration of the *FM-6*. Several methods of obtaining a "zero beat" indication are available. There is a meter in the receiver which can be used, the beat may be checked audibly, or an oscilloscope presentation may be used.

A standard signal generator manufactured by Marconi is included in the package. It is used for servicing and as a means of measuring the sensitivity of the receivers. The remaining items of equipment in the test bay consist of several attenuators and coaxial switches which are all of a standard design.

In order to allow flexibility in the use of the bay, the inputs and outputs of all equipments are jack-terminated. This will permit the addition of other items of test equipment which may become necessary in the future.

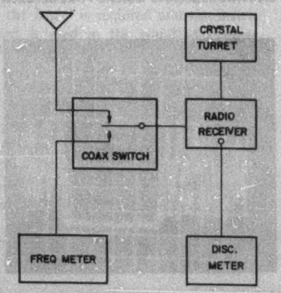


Fig. 4. Block schematic of interconnections for frequency measurement using discriminator

Operations

The frequency measurements are made by using the appropriate receiver and the *FM-6*. The block schematic of the connections is shown in Fig. 4. In operation the input of the receiver is connected through a coaxial switch to an antenna. The switch on the crystal selector is then set to the correct channel. By means of the padder the crystal frequency is adjusted until the discriminator meter reads zero when receiving a signal from the transmitter which is to be measured. The *FM-6* is then connected to the input of the receiver and adjusted until the discriminator meter again reads zero. The frequency may then be read from the *FM-6* after using the proper multiplier.

To avoid the necessity of the operating forces performing any computation, tables similar to Table I have been prepared for all channel frequencies.

Inasmuch as this method is a signal substitution method, the receiver has no influence upon the accuracy of the measurement and it is not necessary to calibrate it in any way.

The first method of deviation measurement to be described has been known for several years but not widely used for several reasons.[2] Since use is made of a receiver discriminator the possibility of the presence of amplitude modulation (AM) cannot be disregarded. With modern high-quality design of limiter and discriminator circuits, the AM suppression of the receiver becomes high enough to permit their use in this way,

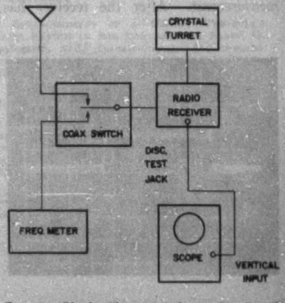


Fig. 5. Block schematic of interconnection for deviation and frequency measurements using oscilloscope

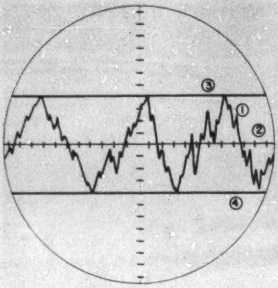


Fig. 6. Typical display of discriminator voltages

1—Modulated r-f signal
2—Unmodulated signal from FM-6 at center frequency
3—Unmodulated signal from FM-6 adjusted to measure upward peak deviation
4—Unmodulated signal from FM-6 adjusted to measure downward peak deviation

if the receiver limiters and discriminator are properly maintained, and a signal level high enough to saturate both limiters can be received.[3] The connections required for this method of deviation measurement are shown in Fig. 5.

Fig. 6 presents a display which is typical of the traces observed when an oscilloscope is connected as in Fig. 5. Line *1* is the trace which represents the modulated signal. Line *2* is the straight line produced by an unmodulated signal from the *FM-6*. Lines *3* and *4* are the traces obtained from the *FM-6* when its frequency has been adjusted to measure the peak deviation.

The initial steps in making the deviation measurement are identical with those taken in making a frequency measurement. After the receiver has been adjusted to give zero discriminator on the unmodulated signal from the transmitter, the controls of the oscilloscope are adjusted to center the straight line on the oscilloscope tube. The input of the receiver is then shifted from the antenna to the *FM-6* and the *FM-6* adjusted to give a straight line at the center of the oscilloscope tube. The input of the receiver is again shifted to the antenna, modulation applied to the transmitter, and the peaks of the voltage display marked on the cathode-ray tube. The input of the receiver is then moved to the *FM-6* and the output of the *FM-6* shifted in frequency until the line from the *FM-6* coincides with the previously marked peaks. At this time the frequency of the *FM-6* is read. The difference between this frequency and the center frequency is the instantaneous deviation of the transmitter.

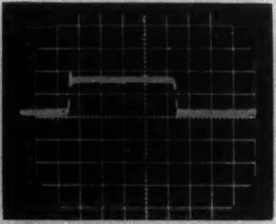


Fig. 7. Center trace from FM-6 adjusted to produce 454.750 mc. Outside trace from transmitter is 1.97 kc low in frequency

A more refined method would be to drive the coaxial switch at a 60-cycle rate, thus displaying the trace from the remote transmitter and the *FM-6* simultaneously. This is the method which is used in the installation being described. Both positive and negative modulation peaks may be measured.

The accuracy of this method was established by making comparisons with measurements made using the Bessel function method of measurement.[4] These measurements indicate that in actual performance on fixed transmitters an accuracy of 7 to 8% can be achieved at the frequencies and deviations being measured.

Figs. 7–11 are oscillograms taken at the test bay. The transmitter being observed was located at Mt. Wilson and operated on an assigned carrier frequency of 454.75 mc. In these figures the trace in the center is from the *FM-6*. The trace at either side is that from the remote transmitter. The vertical distance between the two traces represents the frequency difference between the *FM-6* and the transmitter.

Fig. 7 is the oscillogram obtained when making a frequency measurement. The connections are as shown in Fig. 5. In this specific case the transmitter was 1.97 kc off frequency. Fig. 8 is the oscillogram obtained when the transmitter was adjusted to the correct frequency. This adjustment was confirmed by simultaneous measurement made at the transmitter site. The difference between the two measurements was less than 0.0001%.

Fig. 9 shows the transmitter trace when modulation is applied and Fig. 10 shows the same transmitter combined

Fig. 8. FM-6 and transmitter with transmitter on proper frequency

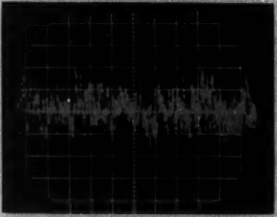


Fig. 9. Trace from transmitter only showing effect of modulation

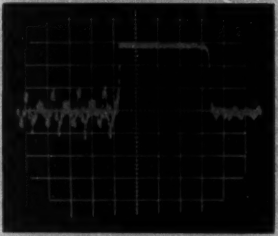


Fig. 10. Outside trace from transmitter showing underdeviation. Center trace from the FM-6 adjusted to indicate 11-kc deviation

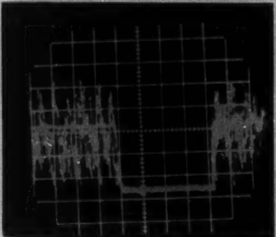


Fig. 11. Same conditions as in Fig. 10. Transmitter has been adjusted to give 11-kc deviation

Vol
Pt.
19

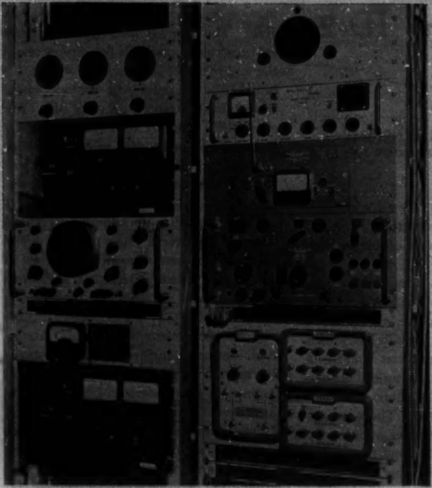


Fig. 12. The complete test bay

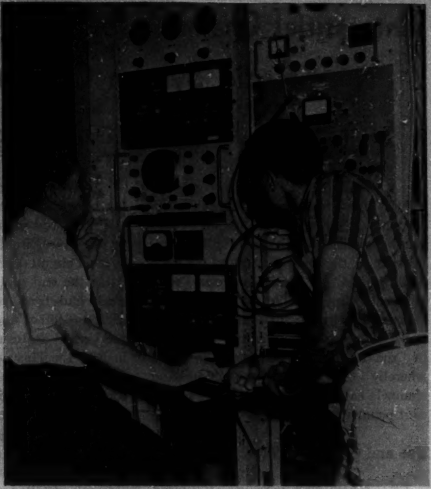


Fig. 13. Test bay in use

with the trace from the *FM-6*. The *FM-6* is displaced 11 kc from the transmitter center frequency. This is the desired deviation. Fig. 11 shows the transmitter after being adjusted to give the desired 11-kc deviation.

A second method of measuring deviation has been provided which makes use of the Gertsch *DM-3* modulation meter and the appropriate one of the receivers mentioned previously. These receivers have been equipped with an additional crystal oscillator manufactured to company specifications which, when mixed with the 290-kc intermediate frequency of the receivers, gives an output at 100 kc. This may be fed directly into the Gertsch *DM-3*. Since this instrument is essentially a 100-kc receiver no modification of the *DM-3* is required and the instrument is used in the standard manner. The frequency of the difference oscillator in the receiver may be shifted as required to satisfy the "frequency set" instructions of the *DM-3*. This equipment has an accuracy of ±5% of the full-scale induction.

The *DM-3* was originally provided to enable measurements to be made on the test transmitters which are located on the same floor as the test bay. To enable frequency and deviation measurements to be made on these transmitters, the r-f transmission line to the antenna is tapped and the branch circuit run through a Dumont attenuator to a jack in the test bay. A patch connection can be made between this jack and the *FM-6* and *DM-3* to allow frequency and deviation measurements to be made on the test transmitters. These measurements are all made in the standard manner.

A third method of measuring deviation of the remote transmitters has recently been suggested and is being given consideration. The proposed method will achieve an accuracy of about ±3% regardless of the frequency band. This method consists of shifting the frequency of the low i-f oscillator. An oscilloscope is connected to the discriminator output as in the first method described. The rest of the procedure is similar to that of the first method with the exception of the point of injection of the variable frequency. If the need for further improvement in accuracy is required, this method will be investigated.

One common defect in a phase-modulated transmitter is the presence of an excessive amount of amplitude modulation of the carrier. This is easily detected by reducing the signal strength of the incoming carrier until it is well below the saturation point of the first limiter. If the oscilloscope is connected to the first limiter jack, any amplitude modulation which may be present is readily detected.

Other tests of the remote and local transmitters will undoubtedly be developed as experience is gained with the equipment.

Fig. 12 is the complete installation, and Fig. 13 shows test bay in operation.

References

1. Establishment of a 450-Mc Vehicular Radio System in the Los Angeles Basin, **G. Applegarth.** *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 79, Nov. 1960, pp. 597–606.

2. Let's Measure Modulation, **Fred Hilton.** *Motorola Newsgram*, Chicago, Ill., Jan.-Feb. 1955.

3. Radiotron Designers Handbook (book), **Langford-Smith.** Wireless Press, Sidney Australia; also reproduced and distributed by Radio Corporation of America, Harrison, N. J., fourth edition, 1952, pp. 1104, 1318, 1292.

4. A Method of Measuring Frequency Deviation, **Murray G. Crosby.** *RCA Review*, Princeton, N. J., vol. 4, Apr 1940, p. 473.

Interstitial Channels for Doubling TD-2 Radio System Capacity

H. E. CURTIS
NONMEMBER AIEE

T. R. D. COLLINS
NONMEMBER AIEE

B. C. JAMISON
ASSOCIATE MEMBER AIEE

THE TD-2 microwave radio relay system at present provides six broad-band channels in each direction of transmission in the frequency band between 3,700 and 4,200 mc (megacycles), each channel being capable of transmitting several hundred voice circuits.[1] This paper describes arrangements and some of the technical problems involved whereby the number of broad-band channels can be doubled with no significant penalty to the existing channels.

Use and Growth of TD-2 Systems

TD-2 routes were installed at a rapid rate after the first *TD-2* system was placed in service between New York and Chicago in September 1950. During the following 8 years there was a large demand for additional intercity television, message, and private line circuits to meet the expanding business requirements. Because the *TD-2* system gives excellent transmission performance and also has economic advantages over other systems, especially where new routes are needed, a large part of the additional Bell System intercity facilities provided during the 8-year period consisted of *TD-2* channels.

Long video network circuits were obtained almost entirely from *TD-2* channels, and side connections to the backbone video networks were usually *TD-2* facilities. These channels were also widely used to obtain intercity voice and private line circuits. By using the type *L* carrier terminal equipment, up to 600-voice circuits were assigned to one *TD-2* radio channel. This technique uses base-band frequencies only up to about 3 mc for the 600-voice circuits, although the base-band capability of the *TD-2* channel is considerably higher. It was found, however, that when the *TD-2* channel was loaded with more than 600-voice circuits, the noise and intermodulation distortion on the voice circuits were often too high to meet circuit objectives.

By the end of 1958 a microwave network had been established which connected together most of the large cities in the United States and Canada, as shown in Fig. 1. This network was comprised of over 30,000 *TD-2* route miles between some 1,300 individual stations scattered throughout the country. These routes provided about 67,000 miles of intercity television circuits and about 20 million telephone circuit miles. The *TD-2* radio channels installed over the approximately 8-year period comprised about 80% of all intercity video and about 35% of all long-distance telephone circuit mileage in the Bell System plant.

Need for Additional TD-2 Channels

Before 1958, many of the existing backbone routes had all six channels in at least one direction of transmission either in service or planned for service; it was becoming increasingly apparent that six radio channels would not be adequate to take care of the wide-band channel needs and that construction of parallel backbone routes would be necessary.

Television requirements for circuits resulted in some routes with more channels in service in one direction of transmission than in the other. These unbalanced requirements, together with other needs for 2-way circuits, resulted in a number of routes soon reaching their capacities in at least one direction of transmission.

By constructing new *TD-2* routes, additional *TD-2* channels could be obtained in some circumstances, and this was done where practicable. Such new routes required new radio towers and buildings to house the equipment and power plants, which proved to be expensive, particularly where the geographical length of the routes had to be increased in order to avoid interference with regard to existing routes.

Other radio systems which would use frequencies in other common carrier bands and which would give high-quality circuits meeting long-distance transmission requirements were not yet available. The new *TH* microwave system is expected to be available for service about the end of 1960, and each individual *TH* broad-band channel will provide about three times the circuit capacity of a *TD-2* channel.[2] The *TH* channels will usually be installed along routes where their high telephone circuit capacity is needed and where higher costs per radio channel will be justified.

The possibility of adding interstitial channels between the regular *TD-2* channels had been considered in view of the unused 20-mc interstitial frequency space available between channels, but a study revealed that it was not practicable with the delay lens antenna and associated rectangular waveguide system.[3] Starting in 1955, the horn-reflector antenna with the associated circular waveguide system became standard for all new installations, and it then became possible to cross-polarize signals between adjacent *TD-2* channels to obtain additional discrimination between adjacent channel signals.[4] By the beginning of 1958, the necessity of providing more circuits along fully loaded *TD-2* routes resulted in the investigation of the feasibility of adding interstitial channels on these routes.

The TD-2 System

The *TD-2* system has been described extensively in the literature, but for the purpose of this paper, certain features will be reviewed briefly.[5] Each base-band channel has a frequency capacity extending from about 30 cycles to 6 to 10 mc per second depending on the length of the system. Frequency modulation is employed to raise each channel to an intermediate frequency (i-f) of 70 mc, from which each is shifted to its proper location in the microwave frequency spectrum. The frequency plan for a typical repeater point is shown in Fig. 2.

Channels are so assigned that there is always a spacing of 40 mc between the carrier of any transmitter operating in a given direction and the carrier of the nearest receiver operating in the same direction. Since each frequency modulation (FM) channel potentially occupies a total band of 20 mc, there is a total of six interstitial bands of 20 mc each in each direction in which no significant quantity of energy is present.

Successful operation of the *TD-2* system requires freedom from interference between microwave channels under all reasonable conditions. The potential interference which does exist arises from radio-frequency (r-f) coupling paths.

Paper 60-1248, recommended by the AIEE Radio Communication Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Chicago, Ill., October 9–14, 1960. Manuscript submitted June 10, 1960; made available for printing August 24, 1960.

H. E. CURTIS and T. R. D. COLLINS are with the Bell Telephone Laboratories, Murray Hill, N. J., and North Andover, Mass., respectively; B. C. JAMISON is with the American Telephone and Telegraph Company, New York, N. Y.

Vol.
Pt. 1
1960

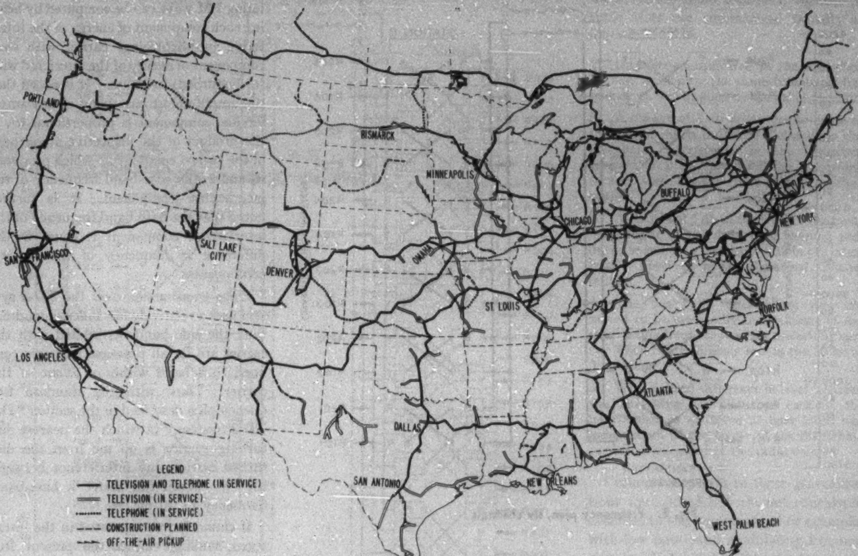


Fig. 1. TD-2 routes, 1958

These couplings may for convenience be separated into two types. The first, called near-end, is due to high-level energy radiated from one antenna on a tower falling into receivers connected to the receiving antenna on the same tower. In the frequency plan of Fig. 2, the significant interferences of this type are 40 mc removed from the center of the band of the disturbed channel. The effect of this r-f interference is minimized by:

1. Transmission loss between the two antennas.
2. Selectivity of the r-f channel-dropping filters and image suppression filters.
3. Discrimination of the i-f amplifiers.

The net loss, including antenna gains between horn-reflector antennas side by side and pointed in the same direction, is in excess of 95 decibels (db) in the absence of foreground reflections. However, measurements at a large number of locations show that the effect of foreground reflections is to distribute the loss over a wide range of values. Fig. 3 shows a cumulative probability distribution curve of 97 measurements of side-to-side coupling loss between horn-reflector antennas. Even 61 db, the poorest value shown, is adequate for 6-channel operation. The r-f filters offer about a 16-db loss at frequencies 40 mc removed from the center of the channel band. The main i-f amplifier, however, provides about 60 db of discrimination at 30 and 110 mc relative to 70 mc.

Since the normal transmitting power as measured at the input to the transmitting waveguide is approximately 0.5 watt, or +27 db below one milliwatt, and the received carrier power at the converter during nonfading conditions is 35 db below one milliwatt for the average path, the level difference between these two points on the same tower is 62 db. However, the sum of the losses above enumerated, together with those in the i-f preamplifier, the transmitter modulator, and the first stage of the transmitting r-f amplifier, insures complete freedom from near-end interference between microwave channels under all normal conditions.

The second type of r-f interference, called far-end, is the result of energy intended for one receiver at a particular point falling into another receiver at the same point. In this case the nearest disturbing channel is 80 mc removed from the center of any particular disturbed channel. For example, at station *B* in Fig. 2 the carrier from station *A* at 4,090 mc may, potentially at least, interfere with the similarly directed channels at 4,170 and 4,010 mc. R-f filters are used which provide about 50 db of loss to frequencies 80 mc removed from the channel-operating frequency; here again the sum of all the circuit losses at side frequencies is sufficient to insure freedom from far-end interference.

Interstitial Channels

Six additional 2-way microwave channels may be derived by utilizing the interstitial bands mentioned above. A frequency plan for such operation is shown in Fig. 4 in which the six additional channels are shown as dashed lines. It will be noted that here adjacent receiving channels are only 20 mc apart, and furthermore that the minimum difference in frequency between transmitters and receivers on the same tower is also 20 mc.

Inasmuch as the total discrimination of the repeater to frequencies 20 mc from mid-band is not great, supplementary means had to be found for improving the discrimination between such closely spaced channels. For far-end interference between similarly directed channels only 20 mc apart, additional discrimination may be obtained by transmitting one set of six channels with vertical polariza-

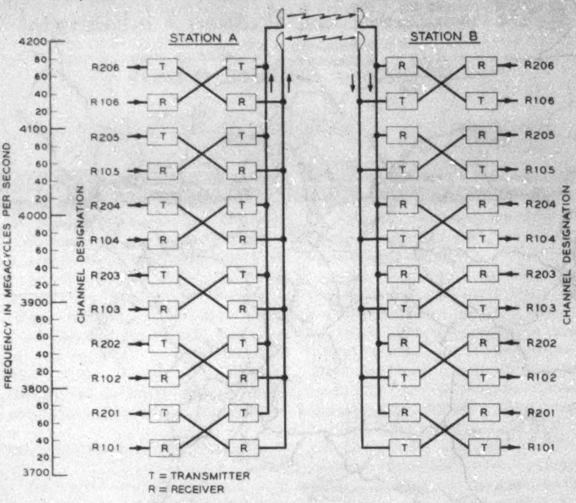


Fig. 2. Frequency plan, six channels

tion, and the other set of six channels with horizontal polarization, as shown symbolically in Fig. 4. The two sets of oppositely polarized channels are separated at the receiving waveguide by means of a suitable network.

It is essential that during fading the amplitude of the vertical component of any particular horizontally polarized carrier shall not become excessive relative to the amplitude of the vertically polarized carrier located 20 mc away, and vice versa. Cross-polarization during fading periods was studied experimentally on a 23-mile path between Murray Hill and Holmdel, N. J. The experimental work was carried out by G. M. Snow, formerly of Bell Telephone Laboratories. This was done by transmitting a vertically polarized carrier at 4,008 mc and a horizontally polarized carrier at 3,980 mc, and observing them with a receiver tuned to each but with both receivers arranged to accept only vertically polarized waves. Fig. 5 is a cumulative distribution curve of the instantaneous difference in level between the two received components obtained during September 1954.

It should be pointed out that by the nature of the recording means, there is an inherent uncertainty of ±5 db in the observed values of discrimination. Fig. 5 shows that during this period the cross-polarization discrimination was poorer than 20±5 db for only about 0.002% of the time.

While the use of cross-polarization materially improves the interference from far-end couplings, it does not appreciably increase the side-to-side coupling loss between antennas; consequently, near-end crosstalk becomes an important consideration in interstitial channel operation.

Interchannel Interference, Theory

In an ideal FM system, the base-band interference due to a relatively weak interfering FM wave can be computed by beating each component of energy of the interfering (or disturbing) carrier with each component of energy of the interfered with (or disturbed) fm wave. It is noted that the amplitude at base band of any interference component is proportional to: 1. the product of the respective amplitudes of the two r-f components which produced it, and 2. the base-band frequency of the interference component. It is further noted that the base-band frequency of the interference component is equal to the difference in frequency of the two r-f components.

Since transmission over the *TD-2* system uses a relatively low index of modulation, the side bands associated with the carrier are for all practical purposes confined to a band within ±10 mc of the carrier. Thus, with the standard frequency plan described in the section "The *TD-2* System," in which the nearest disturbing carrier is 40 mc from the disturbed carrier, any interference between the two would fall above a base-band frequency of 20 mc.

If channels were operated in the interstices available under the present frequency plan, potential interference would extend from zero base-band frequency upward, having maximum power at 20 mc. A quantitative analysis using this theory shows that, for the types of base-band signals normally transmitted over the *TD-2* system, the interference falling in the useful base-band frequency range of the *TD-2* system would be expected to be negligible if the disturbing carrier at the FM detector is always at least 6 to 10 db below the disturbed carrier.

Preliminary laboratory studies and pub-

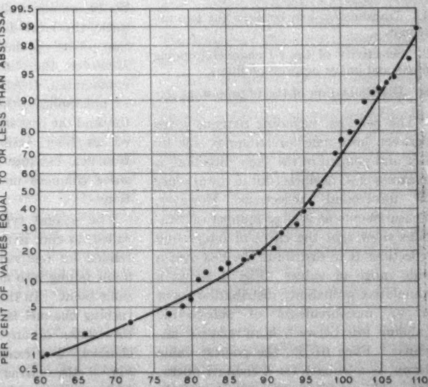


Fig. 3. Side-to-side antenna - coupling loss

Vol.
Pt.
196

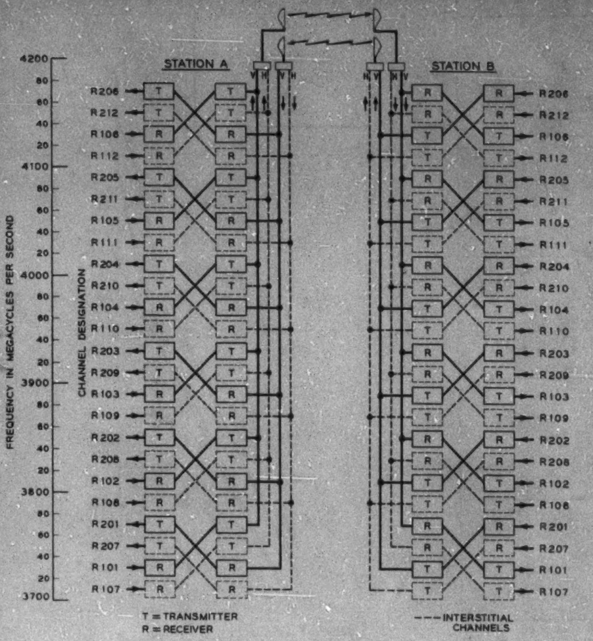


Fig. 4. Frequency plan, 12 channels

lished information indicate that the interference at baseband was far higher than expected, and that it did not obey the laws indicated by ideal, simple fm theory.[6] For this reason a detailed investigation to establish the feasibility of interstitial operation was implemented in the laboratory and continued in the field.

Laboratory Study

Some of the characteristics of direct adjacent channel interference between *TD-2* channels 20 mc apart had been studied briefly in the laboratory as early as 1952. The desire to operate the *TD-2* system as a 12-channel system making use of this spacing made it necessary to extend the previous work and to produce an engineering plan for successful operation of the 12-channel system.

A typical laboratory circuit arrangement for testing such a proposal is shown in Fig. 6 in block schematic form. The upper left-hand part of the figure shows the disturbing channel, operating at 3,830 mc and modulated by a signal from a baseband signal generator. The equipment in the lower left section provides the 3,810-mc carrier which represents the disturbed channel.

In the early part of the studies this carrier was unmodulated, as indicated in the figure. The two microwave carriers are combined in the hybrid junction and delivered to the radio receiver which is tuned to the carrier frequency of the disturbed channel. The 70-mc output of the radio receiver is passed to the fm receiver, which is in turn followed by baseband measuring equipment of appropriate type. The variable attenuator *AT1* is used to adjust the 3,810-mc carrier power at the input to the radio receiver to any desired value. Similarly, *AT2* is used to adjust the received power of the 3,830-mc disturbing carrier. The difference between the decibel settings of the two attenuators represents the cross-polarization discrimination, which proved to be the significant parameter in many of the tests.

Tests were made using single-frequency base-band tones to modulate the disturbing channel, while interference in the disturbed channel was observed at the output of the FM receiver connected to the 3,810-mc radio receiver. Results of these tests are summarized broadly as follows:

1. Interference appeared in the disturbed channel at exactly the same base-band frequency as was applied to the other channel.
2. A change of 1 db in the carrier ratio of the two channels produced a change in the base-band interference in the disturbed channel of 2 db, provided the disturbing carrier was weaker than the disturbed carrier.
3. A change of 1 db in the level of the base-band input to the disturbing channel produced in general a change of 1 db in the interference observed in the disturbed channel output.
4. The base-band interference was observed to be essentially independent of the modulation frequency on the disturbing carrier when this frequency was in the 50-kc to 5-mc range.
5. With a fixed difference in level between the disturbing and disturbed carriers, the interference as observed at base band was independent of the power of the disturbing carrier at the input of the radio receiver.

Certain exceptions to these generalizations were noted but are not considered essential to the argument. For example, with low base-band modulating frequencies, i.e., 50 kc, a change in the frequency deviation from ±4 to ±0.4 mc produced a change in interference in the disturbed channel of 26 rather than 20 db.

Fig. 7 shows some of the characteristic relations between the FM carrier ratios and the base-band interference. The abscissa is the ratio in decibels of the two carrier powers as measured in the receiving waveguide. The ordinate is the observed base-band interference in decibels on an equal level crosstalk scale.* Over a considerable range of carrier ratios the slope of the curves is essentially 2 to 1. However, as the carrier ratio begins to approach unity, the slope is considerably greater. Since a working system will not be satisfactory when the interference is large, the region where the disturbing carrier is stronger than the disturbed carrier is only of academic interest.

While the shapes of the curves were almost the same over the region of interest, the position of the curves with respect to the carrier ratio depended upon the discrimination against the disturbing channel that was provided by the frequency characteristic of the receiver in the disturbed channel. The experimental work indicated that the crosstalk mechanism which is characteristic of adjacent channel interference existed in the limiter of the

---

* On this scale the number −40 means that the measured interference at base band is 40 db below the value of the base-band tone which would be measured at the output of an FM receiver on the disturbing channel.

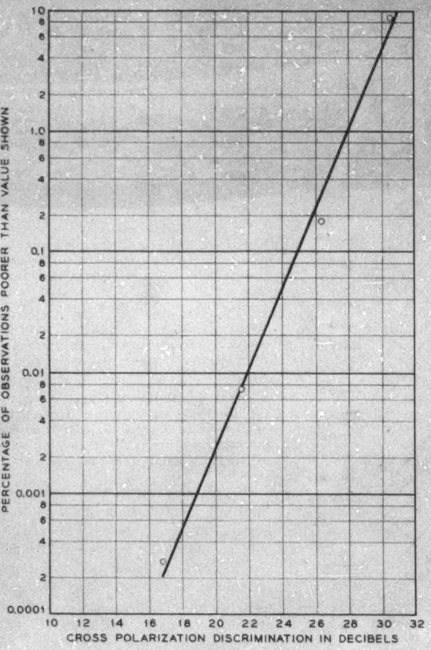


Fig. 5. Cross-polarization discrimination

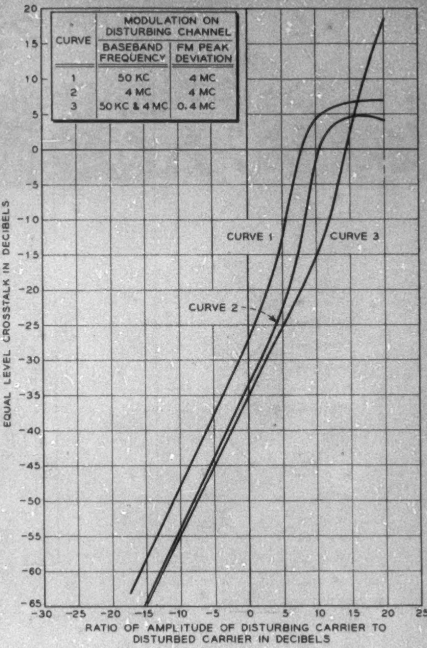


Fig. 7. Crosstalk between microwave channels, laboratory measurements

fm receiver, and also at any point in the system where there was a pronounced tendency to compress the signal.

Such a mechanism has been suggested in an unpublished memorandum by W. R. Bennett of the Bell Telephone Laboratories. This mathematical work shows that when an unmodulated FM carrier and a weaker sinusoidally modulated FM wave are passed through an ideal clipper limiter, distortion is introduced on the stronger FM carrier, some of which appears at base-band frequencies as interference having the same frequency as that of the modulation on the weaker carrier. The amplitude of this interference varies in two ways: as the square of the ratio of the amplitude of the two carriers, and as the square of the ratio of the voltage at which the limiter clips to the signal voltage impressed upon it.

The next series of experiments was undertaken to determine how the frequency discrimination at the edges of the channel pass band might be changed to hold the interchannel interference to tolerable limits in the presence of a limited amount of cross-polarization discrimination. It was found that sufficient improvement could be obtained if the frequency discrimination of the disturbed channel was increased in the region of the disturbing carrier and its signal side bands. Since the strength of these side bands usually decreases as their frequency spacing from the carrier increases beyond

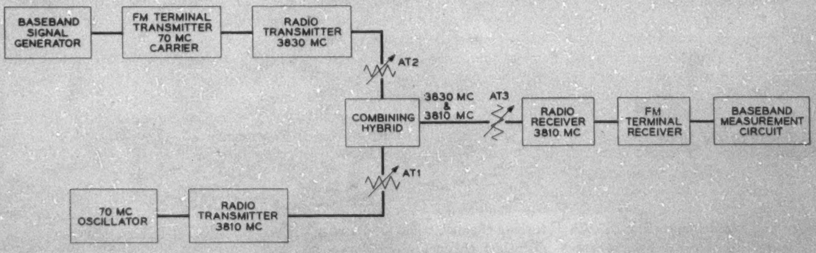

BASEBAND SIGNAL GENERATOR
FM TERMINAL TRANSMITTER 70 MC CARRIER
RADIO TRANSMITTER 3830 MC
AT2
COMBINING HYBRID
3830 MC & 3810 MC
AT3
RADIO RECEIVER 3810 MC
FM TERMINAL RECEIVER
BASEBAND MEASUREMENT CIRCUIT
AT1
70 MC OSCILLATOR
RADIO TRANSMITTER 3810 MC

Fig. 6. Laboratory test arrangement

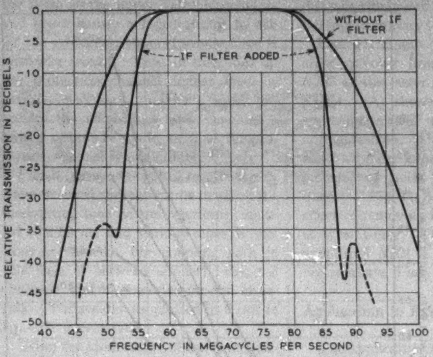


Fig. 8 (above). Radio receiver gain-frequency characteristics

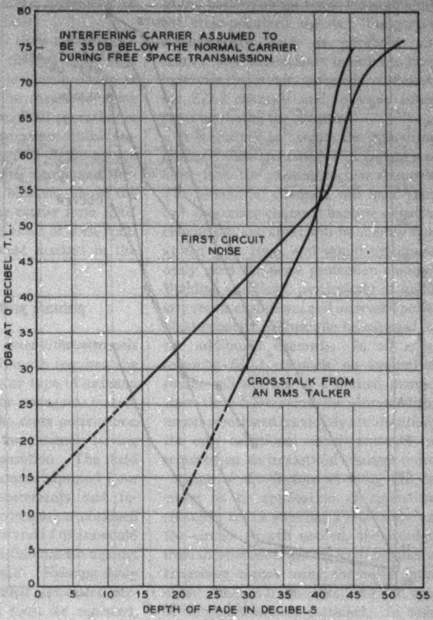


Fig. 9 (right). Expected noise and crosstalk versus fade

the peak frequency deviation, and since the peak deviation is 4 mc in the *TD-2* system, an i-f filter which would increase the discrimination markedly in the range 16 to 24 mc from the disturbed carrier, and somewhat less strongly in the region 10 to 16 mc removed, should be effective. Laboratory models of such a filter were made. The filter was placed in the receiver at a point between the i-f preamplifier and the i-f main amplifier. This arrangement was also tested in field trials, and both laboratory and field results were encouraging. Fig. 8 shows the shape of the pass band of a receiver before and after adding an experimental i-f filter. The disturbed carrier will be located at 70 mc and the disturbing signal will be centered either at 50 or 90 mc. In many cases disturbing signals will be located on both sides.

Since the *TD-2* system normally operates with compression in the final stage of the microwave transmitting amplifier, this i-f filter, for the reason described above, must be applied at each repeater point, and must be located between the point of exposure and the transmitting r-f amplifier.

Another part of the general interference problem involves the condition in which the carrier of the disturbing channel has been severely faded or actually lost at some preceding section. In the latter case, the gain of the system is sufficient to amplify the noise originating in the converters until the transmitter output consists of a broad band of noise whose total power is substantially that of the carrier it replaces. In this situation the energy density at the edge of the band is greater than it is when a carrier with its normal side-band distribution is present. The tendency to interfere or "spill over" into the adjacent channel is greatly increased.

The automatic switching system for *TD-2* operates when the high-frequency noise in a broad-band channel reaches a predetermined level as a result of a fade or loss of carrier in a preceding section of the system.[7] If spillover from a failed interstitial channel occurs, a nominally good channel may be impaired and an unnecessary switch will ensue. The filter is not adequate to cope with this situation, and consequently during such periods the repeater gain must be reduced.

The elements of the engineering plan for interstitial operation may be found in the preceding discussion. A new i-f filter and means for reducing the repeater gain during deep fades must be provided at each repeater on an interstitial route. Cross-polarization discrimination is available in the form of a polarization-sensitive 4-kilomegacycle (kmc) microwave network which connects the rectangular waveguide into the round waveguide leading to the antenna. This network had been developed as part of the equipment necessary to permit systems operating at 4, 6, and 11 kmc to be simultaneously connected to one horn-reflector antenna.[8]

Field tests of various aspects of the engineering plan were carried out from time to time. Before discussing these tests, it seems worth while to discuss the objectives against which the results of the field tests were evaluated.

Objectives

The objective for satisfactory interstitial operation is based on the philosophy that the speech crosstalk from an adjacent channel during a fade should be well masked by the first circuit noise simultaneously present. For system engineering purposes, this objective must be interpreted in terms of a required ratio of disturbed carrier to disturbing carrier.

In the case of near-end crosstalk as defined above, the desired received carrier is subject to normal fading, whereas the disturbing carrier is not. For this condition it has been determined experimentally that without the special i-f filter, if the disturbing carrier is 35 db or more below the disturbed carrier during conditions of free space transmission, then during fading, the base-band interference from the interstitial channel into the disturbed channel will always be below the noise in the disturbed channel.

In the case of far-end crosstalk, fading of the disturbing carrier is correlated to a

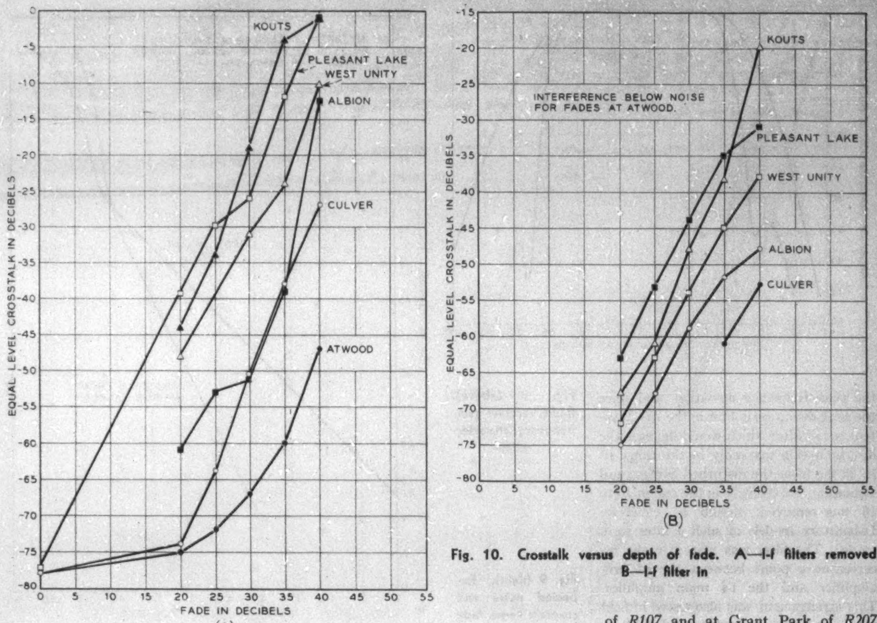


Fig. 10. Crosstalk versus depth of fade. A—I-f filters removed. B—I-f filter in

degree with fading of the disturbed carrier, and this in turn leads to an objective somewhat less than 35 db. The limited amount of data at hand indicates that for this case a ratio of about 20 db is adequate. If the fading on the two channels is uncorrelated, the objective will be 35 db for far-end as well as for near-end crosstalk.

Fig. 9, taken from reference 9, shows how the fluctuation noise and speech crosstalk contributions of one repeater may be expected to increase with depth of fade. Talker volumes are distributed over a wide range, and the volume assumed here is sometimes referred to as the volume of the "rms" talker; this is approximately 10 db below the loudest talker. At a 35-db fade, the point at which service is normally transferred to a nonfaded channel, the speech crosstalk is 7 db below the noise power.

It is also convenient for some purposes to specify the above objective in terms of equal level crosstalk at base-band frequencies. The conversion can readily be made by means of the experimental relationship shown in Fig. 7. Thus, if the 35-db objective were met, then under free space conditions the equal level crosstalk would be completely negligible. For a 35-db fade, the carrier ratio would be 0 db and the corresponding equal level crosstalk objective would be −35 db.

Field Tests

In the six repeater sections between West Unity, Ohio, and Grant Park, Ill., interstitial channels had already been installed, thus providing a suitable location for feasibility field tests. Channels *R201* and *R207* in the east-to-west direction and *R101* and *R107* in the opposite direction (see Fig. 4 for channel nomenclature and frequencies) were used for the tests since *R101* could interfere with *R107* and, produce far-end crosstalk and similarly, for example, it could couple into *R207* and produce near-end crosstalk. Both crosstalk paths involve a spacing of 20 mc. The following basic tests were among those made:

Sine Wave Crosstalk

As an example, a sine wave of suitable power was applied to the base-band input of *R201*, and the magnitude of the crosstalk at this same frequency was observed at West Unity at the base-band outputs of *R107* and at Grant Park of *R207*. This was carried out for various values of simulated r-f fades at each repeater section on *R107* and *R207*.

Video Crosstalk

The test detailed in test 1 was repeated using a video signal. In this case the outputs of *R107* and *R207* were examined on a video monitor for evidence of picture crosstalk for simulated fades, as described above.

Noise Spillover

The carrier was removed from one radio channel at the head end of the system, the gain of the subsequent repeaters thereby rising, as described previously, until the repeater output consisted of a broad band of noise. The magnitude of the spillover of this noise into the adjacent channel was observed for various depths of fades.

Fading Margin Loss From Noise Spillover

In a *TD-2* system, one channel out of six is normally set aside as a protection channel. In each operating channel, noise is sensed at a base-band frequency of about 9 mc, and if the noise is excessive a switch is made to the protection channel when available. Noise spillover increases the noise at the upper video frequencies and hence causes a faded circuit to ask for the protection channel sooner than it otherwise would. This, in turn, ties up the protection channel longer than is actually necessary, and as a result increases the amount of time in which a switch cannot be made, because the protection channel is unavailable.

Fig. 10(A) shows typical results of sine wave crosstalk measurements in which the modulating tone on the interfering channel is 3.59 mc and various repeater sections are faded. The coupling path in this case is between similarly directed channels, and it generates far-end crosstalk. The results are plotted in terms of equal level crosstalk.

Using the more severe of the two objectives suggested previously for far-end crosstalk, it was shown that this led to an objective of −35-db for equal-level crosstalk loss for a 35-db fade. It is apparent, then, that three out of the six far-end coupling paths failed to meet this requirement.

The addition of the supplementary i-f filter substantially reduced the magnitude of the crosstalk, as shown in Fig. 10(B). In this case the poorest equal level cross-value for a 35-db fade was −35 db, which meets the stated objective.

Observations showed video crosstalk mixed with the noise for deep fades, but this was eliminated by the application of the supplementary i-f filter.

Figs. 11(A) and (B) show the effect of noise spillover from channel *R201* into *R207* as the latter is artificially faded at Kouts: (A) shows the measured noise spectrum on *R207* with *R201* operating in normal fashion. These curves exhibit satisfactorily the expected 6-db-per-octave increase of noise with base-band frequency. Fig. 11(B), however, shows the effect of removing the carrier from *R201* at West Unity. The effect of noise spillover from *R201* is most marked in the region above 6 mc.

Application to Existing Routes

The application of interstitial channels to an existing *TD-2* route requires the use of the horn-reflector type of antenna and associated circular waveguide system in order to permit the cross polarization of the radio signals for channels having 20-mc frequency separation. The field tests confirmed the results expected from the laboratory measurements and indicated that the discrimination obtained from cross polarization and i-f filters would provide satisfactory discrimination against adjacent channel signals. Existing delay lens antennas with their associated rectangular waveguides must be replaced, and coupling networks must be installed at the bottom of the tower for combining the cross-polarized signals. Temporary arrangements, such as parabolic antennas, may be necessary in order to keep in service the circuits assigned to the existing *TD-2* channels and to avoid interruptions during the cutover.

It is planned to operate the interstitial channels added to a route as a system separate from the existing regular channels. The interstitial channels will have their own protection channels because adjacent channels 20 mc apart will frequently fade at the same time, and would simultaneously need the same protection channel. Building additions, particularly at auxiliary repeater stations, and increased power plant capacity will usually be required for the additional channels. In all other respects *TD-2* channels on interstitial routes will operate in a normal manner since the i-f filter and the gain-reducing means mentioned previously are essentially the only additional circuit arrangement required on an interstitial channel route.

Some of the factors affecting the decision as to application of interstitial channels to an existing *TD-2* route are the circuit growth needed, the freedom from interferences arising from parallel or branching routes using the same interstitial channel frequencies, and the type of antennas already installed. In some cases there may be interference from

1. 79
1
60

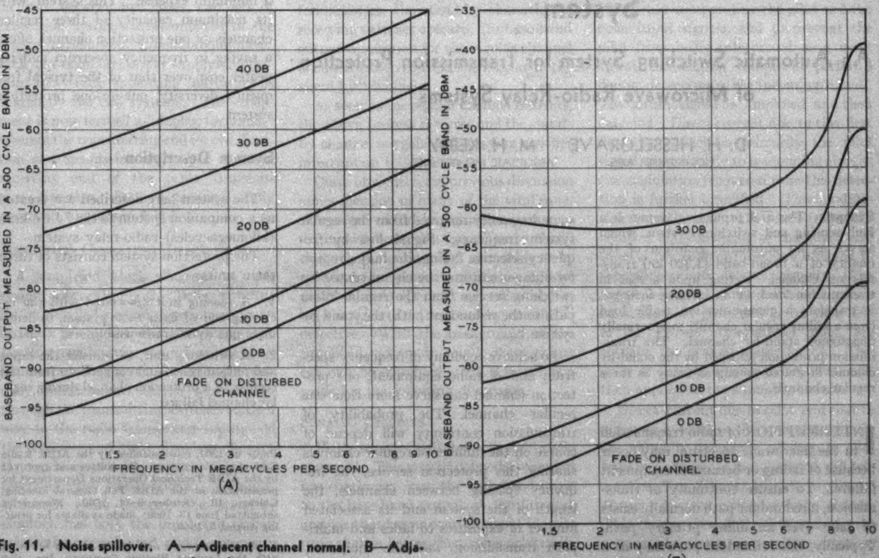


Fig. 11. Noise spillover. A—Adjacent channel normal. B—Adjacent channel open

other radio routes such as side legs or converging routes which might limit the number of interstitial channels which could be added.[9]

TD-2 routes having horn-reflector antennas are particularly attractive candidates for interstitial channel applications from an economic standpoint. The cost of replacing the delay lens antennas and rectangular waveguides by horn-reflector antennas and circular waveguides, and at the same time maintaining existing service, is substantial.

In general, the same factors apply for side-leg routes as for main through routes. However, in side-leg route cases it will be necessary to make sure that the addition of interstitial channels on the side legs will not restrict the number of such channels which could be added on main routes because of possible frequency interferences.

Conclusions

It appears that widespread use of *TD-2* interstitial channels is possible since the results of the field trial indicate that the interference between nearby and branching *TD-2* routes would in general be sufficiently low to permit the use of the same channel frequencies on different routes. In a few cases, however, it is expected that interroute interference might be severe enough to prevent using these channels on more than one route unless remedial measures are taken, such as the construction of new route for short distances.

The results from the field tests and from experience with interstitial channels placed in service in the Chicago-Mishawaka, Indianapolis–Terre Haute, and Troy Hill–Hopedale sections subsequent to the field trial, indicate that both the regular and interstitial channels have transmission characteristics practically the same as those of *TD-2* channels on noninterstitial routes. In view of these satisfactory results, plans are being made for the installation of interstitial channels on *TD-2* routes in many portions of the United States. It appears that nearly all the main *TD-2* routes will have interstitial channels added during the next few years.

References

1. A BROAD-BAND TRANSCONTINENTAL RADIO RELAY SYSTEM, **T. J. Griesser, A. C. Peterson.** *Electrical Engineering*, vol. 70, Sept. 1951, pp. 810–15.
2. 6,000-MEGACYCLE-PER-SECOND RADIO RELAY SYSTEM FOR BROAD-BAND LONG-HAUL SERVICE IN THE BELL SYSTEM, **M. B. McDavitt.** *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 76, 1957 (Jan. 1958 section), pp. 715–22.
3. ANTENNAS FOR THE TD-2, **A. H. Lince.** *Bell Laboratories Record*, New York, N. Y., vol. 30, Feb. 1952, pp. 49–55.
4. A NEW BROAD-BAND MICROWAVE ANTENNA SYSTEM, **R. W. Friis, A. S. May.** *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 77, Mar. 1958, pp. 97–100.
5. THE TD-2 MICROWAVE RADIO RELAY SYSTEM, **A. A. Roetken, K. D. Smith, R. W. Friis.** *Bell System Technical Journal*, New York, N. Y., vol. 30, pt. 2, Oct. 1951, pp. 1041–77.
6. DISTORTION IN FREQUENCY DIVISION MULTIPLEX FM SYSTEM DUE TO AN INTERFERING CARRIER, **R. G. Medhurst, Mrs. E. M. Hicks, W. Grossett.** *Proceedings*, Institution of Electrical Engineers, London, England, vol. 105, pt. B, May 1958, pp. 282–92.
7. PROTECTION OF SERVICE IN THE TD-2 RADIO RELAY SYSTEM BY AUTOMATIC CHANNEL SWITCHING, **I. Welber, H. W. Evans, G. A. Pullis.** *Bell System Technical Journal*, vol. 34, May 1955, pp. 473–510.
8. A NETWORK FOR COMBINING RADIO SYSTEMS AT 4, 6, AND 11 KMC, **E. T. Harkless.** *Ibid.*, vol. 38, Sept. 1959, pp. 1253–67.
9. RADIO FREQUENCY INTERFERENCE CONSIDERATIONS IN THE TD-2 RADIO RELAY SYSTEM, **H. E. Curtis.** *Ibid.*, vol. 39, Mar. 1960, pp. 369–87.

The 57A Microwave Path Protection System

An Automatic Switching System for Transmission Protection of Microwave Radio-Relay Systems

D. H. HESSELGRAVE
NONMEMBER AIEE

M. H. KEBBY
ASSOCIATE MEMBER AIEE

Synopsis: The 57*A* protection system is a fault-sensing and switching system, which serves as an automatic guardian of the quality of a broad-band (1,300 kc) transmission channel. Acting upon a loss in transmission level, or an increase in noise, it transfers a communication traffic load from a failing regular channel to a normally unassigned stand-by channel. The transmission protection afforded by the stand-by channel is shared among as many as three regular channels.

INTERRUPTION of radio transmission in the microwave spectrum may occur because of fading or because of equipment failure. To ensure continuity of transmission, a redundant path normally exists to take over an ailing primary path. Typically, this redundant path may be a parallel radio-relay system, operating at some frequency removed from the regular system frequency. Signal loss by frequency-selective fading, the major reason for outage, can now be circumvented by switching service from the regular radio path to the redundant path, the stand-by system.

To achieve economy of frequency spectrum and of radio equipment, one protection channel can serve more than one regular channel. The probability of transmission continuity will depend of course on the number of regular channels sharing this protection service, the frequency spacing between channels, the length of the system and its associated number of exposures to fades and multipath transmission, and the equipment reliability.

The *57A* microwave path protection system permits an orderly expansion of microwave path transmission capacity at a minimum expense. This system, with its maximum capacity of three regular channels for one protection channel, offers a saving in frequency spectrum and repeater cost over that of the typical frequency diversity, one-for-one protection system.

System Description

The system here described was created as a companion system to the *74A* 6-kmc (kilomegacycles) radio-relay system.

The protection system consists of three main units:

1. A sensing unit associated with the receiving end of each radio system, to determine quality of transmission.
2. A switching unit, to transfer the input and output base-band connections from the regular to a protection channel during regular channel failure.

Paper 60-1207, recommended by the AIEE Radio Communication Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Chicago, Ill., October 9–14, 1960. Manuscript submitted June 13, 1960; made available for printing August 5, 1960.

D. H. HESSELGRAVE and M. H. KEBBY are both with the Lenkurt Electric Company, Inc., San Carlos, Calif.

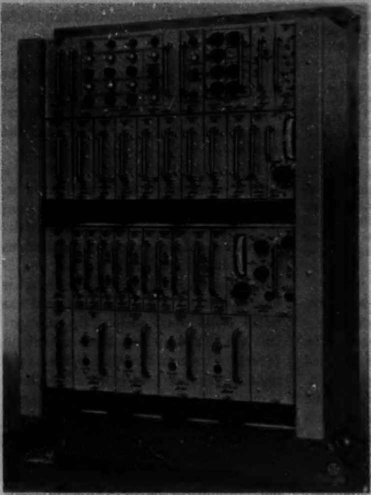


Fig. 1. Fully equipped duplex terminal (transmitting and receiving ends) of a 57A protection system

3. A supervisory control path between the sensing unit at the receiving end and the switch elements at the transmitting end of the radio system.

This equipment protects a unilateral transmission path. Since most of the practical applications are on a bilateral transmission path, such as the telephone facilities between two cities, two sets of protection systems are combined in the typical installation. Common equipment, such as a regulated power supply and certain relay-rack mounting hardware is shared by the two systems. The equipment is now termed a "duplex terminal," wherein the transmitting end for one direction of transmission is grouped with the receiving end of the other transmission direction. Such a terminal, fully equipped, appears in Fig. 1.

SIMPLIFIED OPERATING DESCRIPTION

A simplified block diagram of this system appears in Fig. 2. Before entering into a detailed system description, let us consider briefly how the system operates. Under normal service conditions, the base-band inputs for the regular channels, typically, multichannel carrier systems, go through the transmitting switches on their way to the radio transmitter input. At the receiving end, these signals go through the receiving switches on their way to the receiving drop terminations. Normally, a fourth radio system, the stand-by channel, has both the input and output connections terminated, and is not carrying any service.

When a sensing circuit recognizes a failure of its associated regular channel, a command signal is sent to the transmitting end; the base-band input bridges the stand-by channel, and the base-band signal now rides on two radio paths. When the receiving end of the switching system recognizes regular channel information on the stand-by channel, the receiving switches operate; the base-band output connection for the regular channel is now transferred from the regular to the stand-by channel.

As soon as the regular channel recovers, the above process reverses and the stand-by channel is again idle, waiting another interruption to service.

Quite obviously, the previous discussion omits mention of many of the vital functions required for the protection system to distinguish between a "good" and a "bad" channel, to decide which channel is to be switched, to prevent multiple or false switches, and to prevent any switches during the time that the stand-by channel is defective. With the background of the simplified system in mind, these points can now be examined. Fig. 3 serves as a guide for the following discussion.

SENSING CIRCUITS

First, what constitutes a "good" radio channel? This may be defined as one that has continuity from the base-band input to the base-band output at the far end, and one that has an over-all signal-to-noise ratio that is better than some arbitrarily chosen figure.

To establish continuity, the protection system monitors a pilot which is injected at the transmitting end of the system. This pilot lies above the commercially usable toll telephone base-band spectrum, at a frequency of about 1,600 kc. Its transmission level is a compromise between one that would be so high as to load the radio system, and one that is so low that the superimposed modulating tones could not be recovered in the presence of system noise. This compromise level is about 12 db (decibel) below that of a single v-f (voice-frequency) channel test tone, as applied from a 264-channel carrier system.

So that the receiving end of the system can determine if the stand-by channel is connected to a regular channel input or is merely idle, only the regular channel pilots carry a superimposed 4-kc modulation; the presence of this modulating tone at the stand-by channel sensing equipment signifies that a transmitting-end switch has been completed.

In addition to the pilot, the protection system also monitors the noise lying in a broad slot surrounding the pilot. Although this part of the radio base-band lies above the commercially used spectrum, the noise characteristic is closely related to that within the commercial band; furthermore, the transmission roll-off is not of sufficient magnitude appreciably to alter the received signal level over normal length radio systems.

For ease in separating the pilot and the noise input signals, and to prevent the pilot from being interpreted as a noise signal, the sensing equipment performs double-detection on the incoming signals. The entire signal is amplified and then detected. Direct current due to this first detection, which is primarily the pilot signal, goes directly to a summing circuit; the modulation recovered from this detection is further amplified. This modulation must be split; the lower-frequency components, consisting of the guard and supervisory control tones, are directed to the appropriate control circuits. The higher-frequency components, lying in a band of about 30 to 100 kc on each side of the 1.57-mc (megacycle) pilot, are actually the system noise signals. This noise is further amplified, rectified, and then applied to the summing circuit.

Direct current due to pilot and that to noise, as applied to the summing circuit, are mutually opposing. A drop in pilot level or an increase in noise level will trip a "slicer," or Schmitt trigger, to give an abrupt, discrete control-voltage change. This change sets off the first event in the sequence leading to a complete protection switch.

The noise signals previously mentioned result either from radio transmitter failure or atmospheric fading. In either event, the gain of the radio receiver will rise seeking an r-f (radio-frequency) signal. In doing this, it of course increases the base-band output level of the Johnson noise arising in the front end of the radio receiver. The system may be operated to switch at noise levels ranging between 20 dba (decibels adjusted) minimum to about 55 dba maximum (as measured with a telephone-type noise measuring test set on the v-f drop referred to 0-db transmission level point). The pilot threshold is set so that a 3- to 4-db loss in pilot level will initiate a switch. In both the pilot and noise recognition circuits a fair degree of inertia or hysteresis must exist to prevent jittering between a switch and no-switch condition right at the threshold point.

First Phase of Switch Sequence: Request for Switch Protection

A discrete voltage resulting from either low pilot or high noise is applied to the input of the receiving-end logic circuit. This circuit has three distinct functions as follows:

1. It must initiate a transmitting-end switch request, but only if no other regular channel has already called for a protection switch.

2. It must order the completion of a receiving-end switch (the last event in the switch sequence), but only after the stand-by channel has bridged onto the regular channel at the transmitting end.

3. In the event of simultaneous failure of more than one channel and subsequent return to normal of a switched channel, it must delay the switch sequence to the second channel until the entire switch sequence from the first channel has been restored to normal.

At this stage of the switch sequence, only the first function is of importance.

The d-c output current from each of the radio channel sensing circuits is assigned to an input of the receiving logic circuit. These four inputs feed reverse-biased d-c amplifiers which, for the regular channels only, are interconnected. A request for switch protection on any one of the inputs will drive the associated amplifier into saturation; the output of this amplifier, through the interconnection, immediately clamps the other two regular channel amplifiers in a cut-off condition. Because of the electrical "over-center," bistable, nature of these interconnected amplifiers, only one of the three can ever conduct at one time.

No order of preference exists for the regular inputs. They operate on a "first come, first served" basis, with the fastest sensing chain seizing the stand-by channel service.

The conducting amplifier turns on a regular channel control tone oscillator, operating in the voice frequency region; its assigned frequency tells the transmitting end of the switch section which channel is calling for a switch.

The first phase of the switch sequence is now complete. The regular channel has failed, calling for a switch; the logic circuit has locked up the other channels to prevent a concurrent switch request; the regular channel control tone oscillator is energized and is transmitting a signal back to the transmitting end of the protection system.

The regular channel control tone, accompanied by a guard tone (for monitoring continuity of the control line) proceeds back to the transmitting end of the switch section via a common-carrier facility. This may be a wire line paralleling the radio route, for unidirectional radio

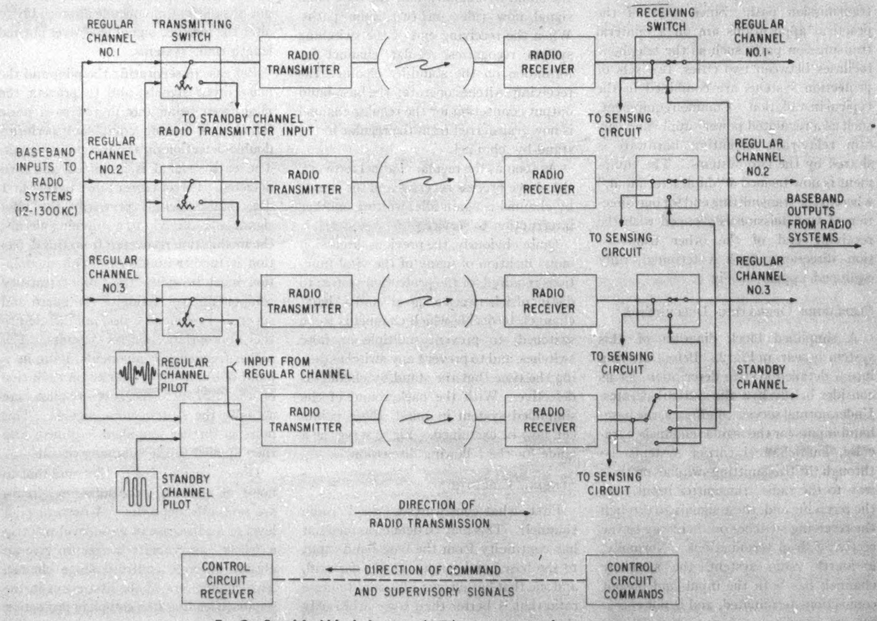

REGULAR CHANNEL NO.1
TRANSMITTING SWITCH
RADIO TRANSMITTER
RADIO RECEIVER
RECEIVING SWITCH
REGULAR CHANNEL NO.1
TO STANDBY CHANNEL RADIO TRANSMITTER INPUT
TO SENSING CIRCUIT
BASEBAND INPUTS TO RADIO SYSTEMS (12-1300KC)
REGULAR CHANNEL NO.2
RADIO TRANSMITTER
RADIO RECEIVER
REGULAR CHANNEL NO.2
BASEBAND OUTPUTS FROM RADIO SYSTEMS
TO SENSING CIRCUIT
REGULAR CHANNEL NO.3
RADIO TRANSMITTER
RADIO RECEIVER
REGULAR CHANNEL NO.3
TO SENSING CIRCUIT
REGULAR CHANNEL PILOT
INPUT FROM REGULAR CHANNEL
STANDBY CHANNEL
RADIO TRANSMITTER
RADIO RECEIVER
STANDBY CHANNEL PILOT
TO SENSING CIRCUIT
DIRECTION OF RADIO TRANSMISSION
CONTROL CIRCUIT RECEIVER
DIRECTION OF COMMAND AND SUPERVISORY SIGNALS
CONTROL CIRCUIT COMMANDS

Fig. 2. Simplified block drawing of 57A microwave path protection system

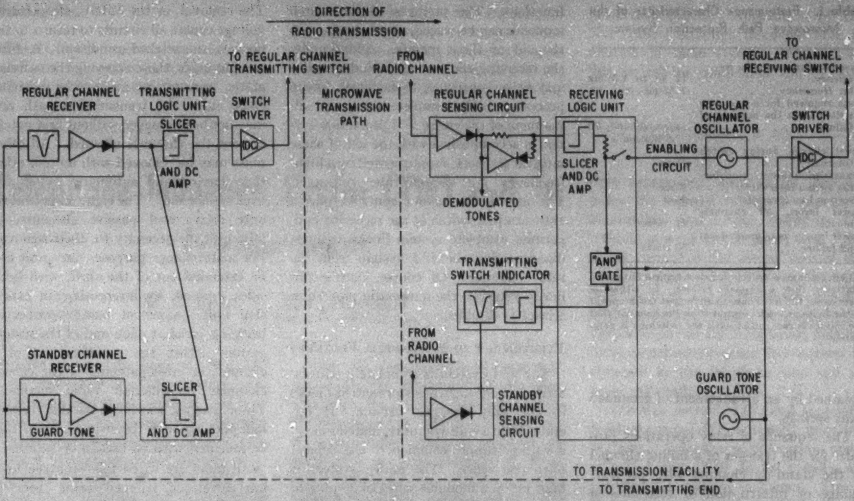

DIRECTION OF
RADIO TRANSMISSION
TO REGULAR CHANNEL TRANSMITTING SWITCH
FROM RADIO CHANNEL
TO REGULAR CHANNEL RECEIVING SWITCH
REGULAR CHANNEL RECEIVER
TRANSMITTING LOGIC UNIT
SLICER
AND DC AMP
SWITCH DRIVER
DC
MICROWAVE TRANSMISSION PATH
REGULAR CHANNEL SENSING CIRCUIT
RECEIVING LOGIC UNIT
SLICER AND DC AMP
ENABLING CIRCUIT
REGULAR CHANNEL OSCILLATOR
SWITCH DRIVER
DC
DEMODULATED TONES
TRANSMITTING SWITCH INDICATOR
"AND" GATE
FROM RADIO CHANNEL
STANDBY CHANNEL RECEIVER
GUARD TONE
SLICER
AND DC AMP
STANDBY CHANNEL SENSING CIRCUIT
GUARD TONE OSCILLATOR
TO TRANSMISSION FACILITY
TO TRANSMITTING END

Fig. 3. Simplified drawing of control circuits for an automatic switch

systems. However, in the general case of the two-way radio system, these control tones ride back as additional modulation on the protection system pilot for the opposite-direction radio system. It is important to recognize here that the *57A* path protection system has two functions, first, that of maintaining system continuity by switching the failed channel to the stand-by channel when necessary, and second, that of serving as a common carrier for the control tones of the complementary, opposite-direction path protection system.

Since there is no more certainty of transmission continuity on this common-carrier path than there is on the forward transmission path, the control tones are multipled to every radio channel for transmission back to the transmitting end. At the far end of the switch section, these tones are demodulated from the separate 1.57-mc pilots, added, and applied to the control tone receiving circuits. Should any return channel fail, its demodulated tone output is removed from the common line to prevent noise or false signal interference.

SECOND PHASE OF SWITCH SEQUENCE: TRANSMITTING-END SWITCH OPERATION

The output of the common-carrier control tone system at the transmitting end of the section is applied to the control circuit receivers. Each of these receivers consists of a band-pass filter, an amplifier, and a rectifier. Just as in the case of the receiving-end sensing circuits, the outputs of these control tone receivers are connected to a logic unit; in this case, it is the transmitting-end logic circuit.

This latter logic circuit has a simpler mission than that of its counterpart at the receiving end. It merely performs the following functions:

1. Permits transmitting-end bridging switches only during the presence of the common-carrier guard tone.
2. Initiates a transmitting-end switch, and concurrently locks up the other two regular channels to prevent any other switch.

Because of the marked similarity in function between the receiving-end and the transmitting-end logic circuits, it was possible to use the receiving-end package, with a few components removed, at the transmitting end.

If the necessary conditions for a transmitting-end switch are fulfilled, a d-c output is delivered from the logic unit to an amplifier which drives a transmitting-end switching relay. This relay bridges the regular channel to the input of the stand-by channel. As a result of this switch, the stand-by pilot is removed and the regular channel pilot appears on the stand-by channel. The regular channel information now leaves the transmitting end of the switching system on two paths, that of its assigned channel and that of the stand-by channel. The second phase of the switching sequence is now complete.

FINAL PHASE OF SWITCH SEQUENCE: RECEIVING-END SWITCH OPERATION

The appearance of regular channel pilot signal, with its accompanying 4-kc modulation, on the stand-by channel is the last "fact" necessary for the receiving logic circuit to order a receiving-end switch. The logical process is: a regular channel has failed and the stand-by channel is bridged to a regular channel at the transmitting end—the failed channel output should now be transferred from the regular to the stand-by channel. Notice that the logic unit does not verify *which* regular channel has been bridged to the stand-by channel. In view of the lock-out features at both the receiving and transmitting ends of the circuit, this refinement and additional safeguard is unnecessary.

The protection switch, as established in the previous sequence, will hold until the regular channel is restored to a "good" condition again. At that time, the entire switch sequence will progressively restore to normal.

SIGNAL SWITCHES

All base-band switching is performed by quick acting, short frame, telephone-type relays with multiple-spring pile-ups. To reduce crosstalk between channels through the relay contacts, additional isolation is

Table I. Performance Characteristics of the Microwave Path Protection System

Base band transmission frequency	12 kc to 1.3 mc
Pilot frequency	1.57 mc
Time required for a complete switch from the moment of failure	approximately 70 milliseconds
Level change during switch interval*	None
Transmission interruption during switch interval*	None
Insertion loss (overall)	4 db
Power input, all channels normal	21 watts
Power input: maximum alarm and failure	50 watts

* This assumes a switch due so a typical fade, or a relatively flow equipment failure. In either of these cases, full level base-band output must appear at the failing channel output from the moment that the failure is recognized until the switching is completed

obtained by an arrangement of grounded relay springs.

The sequence of relay operations provides for the transfer of a failing channel to the stand-by channel with no level change or interruption of transmission during the switching interval.

OPERATING CONTROLS AND INDICATIONS

To avoid any ambiguity, all control functions appear only at the receiving end of the section; each of these operating controls has an associated guard lamp to serve as a reminder of its operation. Two sets of controls are deemed necessary:

1. A control to lock out, or disable, the automatic function of any one, or of all, the channels.
2. A control to switch manually any regular channel to the stand-by channel.

The guard lamps associated with these controls, along with a second set of indicator lamps, tell a complete story on the condition of the protection system. These lamps show which channels have failed, which are locked out; and they indicate that a switch sequence has been completed, or that it failed to complete.

In addition to the receiving-end controls and indications, one set of indicator lamps is also located at the transmitting end of the switch section. These lamps show when a transmitting-end bridging switch has been made to a regular channel. The full set of lamps will indicate the steps in the switch sequence as outlined above, to simplify trouble-shooting.

INTEGRAL TEST EQUIPMENT

A built-in metering panel at the transmitting and at the receiving end of the system reads more than 30 different d-c functions. The progress of the switch sequence can be traced, step by step, with the aid of these meters. Additionally, the receiving-end test set includes a pilot and noise generator, to facilitate lineup procedure. The complete lineup and adjustment of the entire *57A* protection system is accomplished with the use of these integral test sets, supplemented by a high-frequency a-c vacuum-tube voltmeter. The use of the lockout controls, plus a TEST-NORMAL switch at the receiving end, permits complete system lineup without disconnecting the *57A* system from the radio channels. Of course, during this lineup interval, the automatic protection service is disabled.

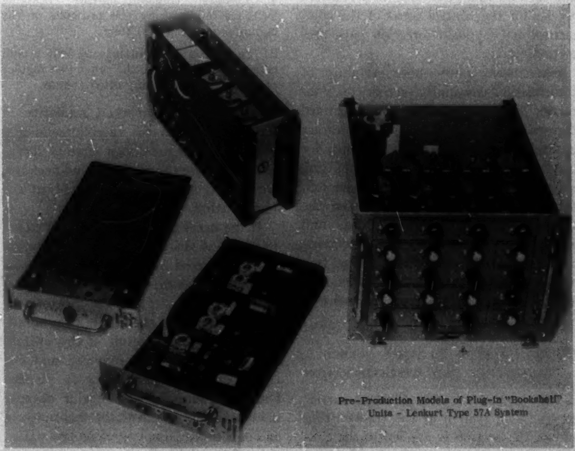
ELECTRICAL AND MECHANICAL FEATURES

The *57A* system is completely transistorized, using 103 transistors and 88 diodes in the fully-equipped version. It operates from a −48-volt unregulated source, a voltage supply common to the typical telephone office. The entire system is built up on a modular series of 30 plug-in units occupying four shelves in a standard relay rack. With no change in common wiring, the system may be expanded from a one-for-one (one stand-by channel to protect one regular channel) up to a one-for-three system, by removing a few plug-in bypass units and plugging in the regular equipped units.

Those units containing circuits common to all channels have an interlock jumper, which removes operating voltage from the system when they are unplugged. The removal, or the failure, of operating voltage causes all circuits to return to the normal, unswitched condition. A total of eight units, those carrying the switches at the transmitting and at the receiving ends, carry the transmission path and may not be unplugged without first patching out the channels affected. All other units may be removed with no risk other than having any automatic switch fall back to normal. The eight units contain only relays and passive elements, to minimize the necessity for their removal. For maintenance purposes, any unit can be extended out of the shelf, with both sides exposed, by interposing an extension unit. A set of parallel-connected bridging jacks at each end of the system permits either the patching out of a channel, or cross-patching of the carrier channels to different radio channels. These patches, which can parallel the normal path through the switch system, can be executed with no breaks or outages.

All units are wired following the Lenkurt-developed stitched-wiring process. This process, adaptable to any circuit consisting of components assembled on a flat dielectric mounting sheet, uses machine-driven staples on a grid pattern to support pigtail components. Wiring between components may either be by hand or on an automated basis. Fig. 4 illustrates the typical unit wiring.

The complete *57A* microwave path protection system, shown in Fig. 1, occupies a total of 14 relay rack spaces, 24.5 inches.



Pre-Production Models of Plug-in "Bookshelf" Units - Lenkurt Type 57A System

Fig. 4. Typical plug-in unit assembly and wiring

Conclusions

The *57A* system permits an orderly expansion of microwave path transmission facilities, at a minimum expense. This system, with its maximum capacity of three regular channels for one protection channel, offers a saving in frequency spectrum and repeater cost over that of the typical frequency diversity, one-for-one protection system. Application of the *57A* system, although discussed here in relation to radio systems, is also feasible for use in connection with a cable or wire line.

Telephone Influence Factor (TIF) and Its Measurement

W. C. BALL
MEMBER AIEE

C. K. POARCH
ASSOCIATE MEMBER AIEE

FROM the earliest days of telephony the evaluation of message circuit noise has been an important problem. Of particular interest has been the evaluation of noise originating from harmonic components present in power system voltages and currents. Magnetic fields set up by power currents, and electric fields set up by power voltages, tend to induce voltages in nearby telephone conductors. Even a small transfer of power at harmonic frequencies in the voice band can cause trouble. For example an input of 400 $\mu\mu$watts (micromicrowatts), or 26 db (decibels) above reference noise, into a telephone set at 1,000 cps (cycles per second) will give an audible tone. To put such a tone on all of the 60 million telephones in the United States at once would require just 24 milliwatts, transmission losses neglected.

If telephone circuits could be perfectly balanced or shielded, the transfer of power at voice frequencies would not be of concern. However, since such perfection is impracticable it is the practice for telephone and power companies to co-ordinate and determine the best engineeering solution to keep power line noise at tolerable levels. For several decades the Joint Subcommittee for Development and Research, of the Edision Electric Institute and Bell Telphone System, has been active in this endeavor.

While the joint co-ordination of systems is a field of considerable scope, the basic yardstick for gaging a particular power circuit as a potential noise source is the telephone influence factor (TIF). This factor is dimensionless and provides an index of the inductive influence of a power system wave shape. It takes into account the relative interfering effect of various frequencies of noise on a telephone circuit (message weighting) and the variations in coupling between power and telephone systems with respect to frequency.

With the introduction of new telephones having improved response characteristics, the relative interfering effects of frequencies in the voice band are changed. Such changes occurred recently with the introduction of the "500-type" telephone set. Thus both the message weighting and the related TIF characteristic have had to be re-examined and brought up to date. The work on the new TIF weighting was done by a Task Force on TIF Revision, set up by the joint subcommittee mentioned above. The revision of the message weighting and the development of a new noise-measuring set incorporating this message weighting was undertaken by the Bell Telephone Laboratories. The set is manufactured by Western Electric Company, Inc., and is coded as the *3A* noise measuring set.

Inductive Co-ordination

The need for inductive co-ordination arises from the basic fact that power system influence is a potential source of telephone noise. If power circuits could be built and operated so as to produce a pure 60-cps sine wave or to prevent any external voice-frequency fields, or if telephone systems were not susceptive to such fields, there would be no need for inductive co-ordination.

There are three basic factors in inductive co-ordination work: the *inductive influence* of the power system, the electric and magnetic *coupling* between power and telephone systems, and the *susceptiveness* of the telephone plant.

The inductive influence of a power system is dependent upon the TIF of the voltage or current and on the voltage or current magnitudes. Voltage TIF affects electric induction which usually is a problem only with open-wire telephone lines. Current TIF is important in cases of magnetic induction. Except for very close separations, such as on joint-use poles, the induced voltage is essentially longitudinal and is due chiefly to the residual current in the power line. Residual current is the vector sum of the phase currents and its TIF may be quite different from that of the phase currents. If the series and shunt unbalances in the telephone circuit are both small, longitudinal induction will cause little current to circulate in the telephone loop and no noise will result.

Although current TIF, and especially residual current TIF, is usually more important in inductive co-ordination, open-circuit voltage TIF limits usually are specified for rotating machines. This is because significant current TIF measurements are not feasible at the factory, and it is assumed that a machine with a low open-circuit voltage TIF will not tend to produce troublesome harmonic currents on a power system.

Susceptiveness, as used in telephony, is essentially a measure of the vulnerability of the telephone plant to outside interference. This is largely a matter of longitudinal balance (as distinguished from return loss balance) and shielding.

TIF, I·T, and Kv·T

The contribution of a given harmonic current (or voltage) to the total noise influence of a disturbing circuit is the product of that current (or voltage) and the appropriate TIF weighting for its frequency. It is designated I·T (or V·T). More often the voltage is expressed in kilovolts and the product is written Kv·T.

Paper 60-1195, recommended by the AIEE Wire Communication Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Chicago, Ill., October 9–14, 1960. Manuscript submitted June 10, 1960; made available for printing August 2, 1960.

W. C. BALL is with the Bell Telephone Laboratories, Murray Hill, N. J., and C. K. POARCH is with the Edison Electric Institute, New York, N. Y. Members of the Task Force on TIF Revision were A. J. Aikens, O. H. Coolidge, M. D. Fagen, H. A. Gauper, Jr., W. W. Holloway, W. R. Hyatt, Jr., C. K. Poarch, L. T. Rosenberg, D. F. Shankle, H. P. St. Clair, T. J. Talley, and M. D. Westrate.

Of the many individuals who have contributed to the development of the new weighting and the new noise measuring set, the authors are particularly indebted to A. J. Aikens, W. T. Cochran, and D. A. Lewinski of the Bell Laboratories, in connection with the 3A set.

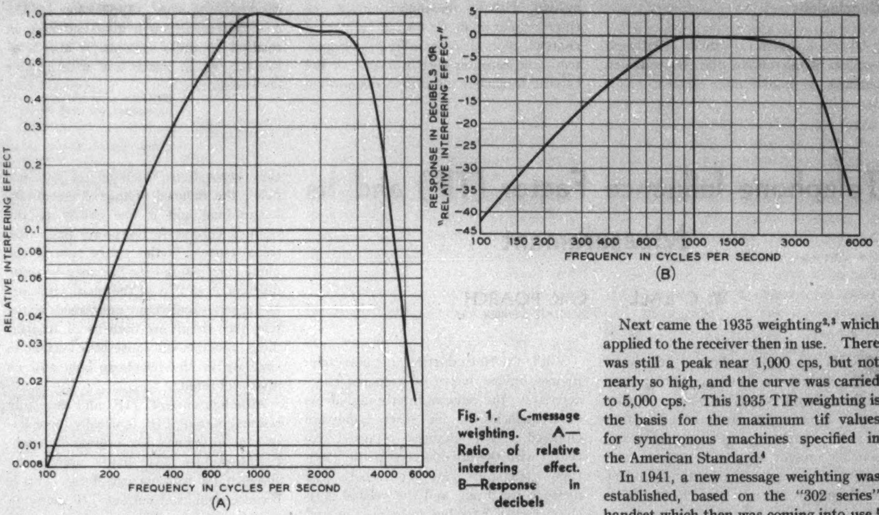


Fig. 1. C-message weighting. A—Ratio of relative interfering effect. B—Response in decibels

The I·T (or Kv·T) of a complete current (or voltage) wave is defined as the square root of the sum of the squares of all of the I·T (or Kv·T) products of the wave's components, including the fundamental.

The current or voltage TIF of the complete wave is the ratio of the I·T or Kv·T value to the total rms value of the wave.

To make TIF a meaningful measure requires that the TIF weighting (TIF_f) account for:

1. The relative subjective effect of frequency f in the message circuit, as heard in the telephone (message weighting).
2. The coupling between the power and telephone circuits, assumed to be directly proportional to f.

Thus

$$TIF = K p_f \frac{f}{1{,}000} \qquad (1)$$

where p_f is the relative interfering effect at f, such that p_f is equal to 1 at 1,000 cps, and K is a constant.

The factor p_f is the ordinate of the message-weighting characteristic. The message weighting characterizes the interfering effect relative to 1,000 cps, as determined by subjective tests. It may be expressed as a decimal fraction of the interfering effect at 1,000 cps or as decibels down from that value. The value at 1,000 cps is therefore 1.0 in the first case and zero db in the second. The newest Bell System noise weighting, called "*C*-message," is shown as a ratio in Fig. 1(A) and in decibels in Fig. 1(B). In equation 1 the ratio is used.

In deriving the *C*-message weighting, two types of tests were conducted using 500-type telephone sets, one in the absence and one in the presence of speech.[1] In the absence of speech a group of observers were asked to adjust the loudness of 14 different frequencies until the sound of each was judged to be equally as annoying as a −59 db below one milliwatt 1,000-cps reference tone. Then, in the presence of the average level of received speech volume, the same observers were asked to adjust the level of each of these frequencies so that the impairment in the received speech was judged to be equivalent to that produced by the 1,000-cps reference tone. Other tests took into account the interfering effects of combinations of single frequencies and of different bands of thermal noise.

TIF Weighting Curves

Figure 2(A) and Table I show comparisons of the several TIF weighting curves which have been used. The first one, published in 1919, applied to the telephone instruments in use at that time. This curve has a high peak near 1,000 cps, indicating a resonance of the telephone receiver at that frequency. It was not carried above 2,100 cps because the telephone of that day was rather insensitive at higher frequencies.

Next came the 1935 weighting[2,3] which applied to the receiver then in use. There was still a peak near 1,000 cps, but not nearly so high, and the curve was carried to 5,000 cps. This 1935 TIF weighting is the basis for the maximum tif values for synchronous machines specified in the American Standard.[4]

In 1941, a new message weighting was established, based on the "302 series" handset which then was coming into use.[5] The resonant peak at 1,000 cps has disappeared but the weightings above 1,400 cps are higher than before, reflecting the improved transmission characteristics of the instrument. Work on a companion TIF weighting was initiated but not completed. One version of a revised TIF weighting under consideration at that time, usually referred to as the "tentative 1941" TIF weighting, has been used to a considerable extent by the Bell System.

The new 1960 TIF curve is based on *C*-message weighting which was derived from tests on the 500-series telephone sets. It is felt that the response of these sets is such that the weighting is largely controlled by the characteristics of the listeners, and that it is unlikely that such changes as may be made in future types of telephones will require a re-evaluation of TIF.

The original scale of values of TIF was related to weighted receiver μamperes per rms ampere in the power line. With each succeeding revision an attempt has been made to adjust the level of the curve (by adjusting the constant factor of the TIF formula) so that readings on typical power circuits or machines would average approximately the same on the new basis as on the preceding one. This generally has made it unnecessary to adopt radically new figures for TIF objectives for apparatus and circuits.

Improvements in the telephone plant over the years have allowed some relaxation in TIF weighting in the lower part of the voice frequency range. This is because the improved transmission has raised the average received speech volume, and because the flatter frequency response of the receiver has improved understandability. Conversely, the TIF weighting in the upper part of the voice frequency range has been raised because of the increased response of the receiver at these higher frequencies.

When a trial TIF weighting characteristic was computed from the *C*-message weighting with $K=6{,}000$, the constant used in the tentative 1941 TIF weighting, the resulting TIF values computed from analyses on many generators and power lines appeared too high, compared with 1935 weightings. The Subcommittee therefore agreed to reduce K to 5,000, and the resulting characteristic has been accepted as the 1960 TIF weighting. With $K=5{,}000$, equation 1 becomes

$$\text{TIF}_f = 5p_f f \qquad (2)$$

The 1960 TIF curve and a table of values for harmonics of 60 cps are shown in Fig. 2(B). The specified TIF of 0.5 at 60 cps (Table I) is nominal and not computed from the *C*-message weighting curve, which has no value specified for 60 cps.

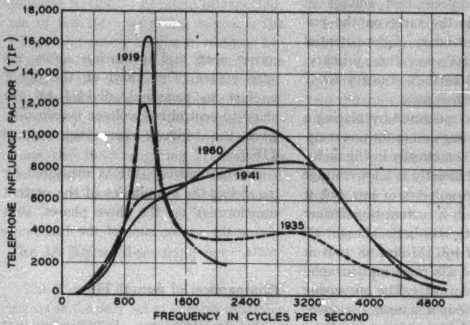
Measurement of 1960 TIF

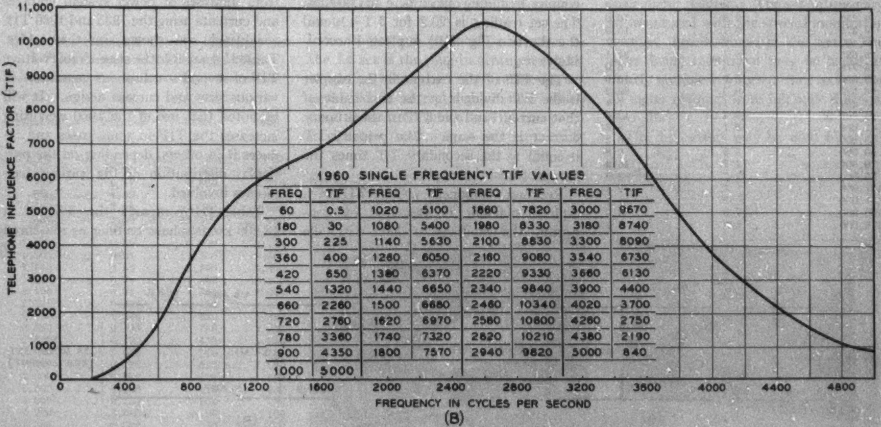
TIF always can be computed by making a harmonic analysis of the voltage or current wave and then applying the individual frequency TIF weightings to the various harmonic components. Such harmonic analyses can be made with any of a number of suitable harmonic analyzers, including the Western Electric *2B* noise-measuring set with the *10A* noise analyzer attachment or the newer Western Electric *4A* noise analyzer.

Less accurate but much more convenient single over-all measurements of voltage or current waves according to the 1960 TIF weighting can be obtained with a Western Electric *3A* noise measuring set and a suitable coupler. The *3A* set includes a *C*-message weighting network. When used with a voltage or current coupler which introduces a multiplier of f, the set readings become proportional to the product of frequency, the relative interfering effect in accordance with the *C*-message weighting, and the magnitude of the voltage or current. Therefore, the measurement is equal to a constant times V·T or I·T. Multiplying the set reading by a suitable constant and dividing by the rms value of the power circuit voltage or current then gives the TIF of the waveform applied to the coupler.

The schematic circuit of a voltage TIF coupling unit as used with the Western Electric *3A* noise-measuring set is shown in Fig. 3(A). The resistor is used to match the input impedance of the set. The accuracy of the readings is directly dependent upon the precision of the 0.01-μfarad capacitor because of its use as a multiplier. The factor for converting a reading to Kv·T is computed by assuming an input of one volt at 1,000 cps. Since the 1960 TIF weighting at 1,000 cps is 5,000, the V·T product of this input is 5,000 and the Kv·T product is 5.

The input impedance of the voltage TIF coupling unit and noise set is $300 - j15{,}900$ ohms at 1,000 cps. With a





1960 SINGLE FREQUENCY TIF VALUES

FREQ	TIF	FREQ	TIF	FREQ	TIF	FREQ	TIF
60	0.5	1020	5100	1860	7820	3000	9670
180	30	1080	5400	1980	8330	3180	8740
300	225	1140	5630	2100	8830	3300	8090
360	400	1260	6050	2160	9080	3540	6730
420	650	1380	6370	2220	9330	3660	6130
540	1320	1440	6650	2340	9840	3900	4400
660	2260	1500	6680	2460	10340	4020	3700
720	2760	1620	6970	2580	10600	4260	2750
780	3360	1740	7320	2820	10210	4380	2190
900	4350	1800	7570	2940	9820	5000	840
1000	5000						

Fig. 2. TIF weighting curves. A—1919–1960. B—1960

Table I. Comparison of Single-Frequency TIF Weightings

Frequency	1935	Tentative, 1941	1960
60	1	0.5	0.5
180	15	61	30
300	205	479	225
420	390	1,190	650
540	1,250	2,090	1,320
660	2,250	3,110	2,260
780	4,080	4,170	3,360
900	7,270	5,160	4,350
1,000	11,000	6,000	5,000
1,020	11,600	6,090	5,100
1,140	11,100	6,310	5,630
1,260	7,920	6,510	6,050
1,380	5,470	6,730	6,370
1,500	4,400	6,900	6,680
1,620	3,900	7,120	6,970
1,740	3,660	7,220	7,320
1,860	3,570	7,380	7,820
1,980	3,500	7,580	8,330
2,100	3,500	7,770	8,830
2,220	3,530	7,940	9,330
2,340	3,580	8,070	9,840
2,460	3,650	8,170	10,340
2,580	3,720	8,220	10,600
2,700	3,800	8,350	10,480
2,820	3,900	8,380	10,210
2,940	3,930	8,450	9,820
3,060	3,950	8,400	9,230
3,180	3,820	8,340	8,740
3,300	3,630	8,060	8,090
3,500	3,090	7,100	7,050
4,000	1,750	3,800	3,760
4,500	1,100	1,300	1,830
5,000	480	423	840

Table II. Generator Open-Circuit Voltage TIF

Kva	TIF	
	1935	1960
2,500	17.2	12.6
6,300	24.8	31.0
9,375	9.6	16.0
9,375	9.1	16.9
9,400	80.0	146.7
15,625	15.7	24.1
15,778	7.9	6.0
23,500	15.7	15.0
23,529	16.3	26.2
35,294	7.1	7.4
35,300	15.4	12.6
41,000	7.7	8.6
43,800	9.1	8.3
50,000	8.9	6.8
70,600	14.7	15.3
75,000	6.7	8.6
107,000	6.7	6.8
107,000	10.5	8.4
125,000	2.8	3.2
176,000	2.9	5.3

1-volt input to the coupler the voltage across the set input is $300/(300-j15,900)=18,900\times10^{-6}$ volts. Reference noise voltage for the new set is 24.5×10^{-6} volts. (This is equivalent to 10^{-12} watts dissipated in 600 ohms, i.e., zero dbrn.) Thus the set will read $20\log_{10}(18,900/24.5)$ or 57.8 dbrn (reference noise in decibels). This is for an input where $Kv{\cdot}T=5$. For $Kv{\cdot}T=1$ the reading will be $57.8-14=43.8$ dbrn. The conversion chart, Fig. 4(A), therefore has an ordinate of 43.8 at $Kv{\cdot}T=1$, and a slope of 20 dbrn per decade.

A similar computation made for any other frequency using the appropriate coupler impedance, TIF weighting and message weighting, will give the same factor of 43.8 dbrn for $Kv{\cdot}T=1$.

The coupler may be connected directly to a low-voltage power secondary or to a potential transformer secondary. The TIF of the voltage is found by dividing the measured Kv·T by the rms voltage in kilovolts. TIF is the same on the primary and the secondary of the potential transformer. Kv·T on the primary, however, is equal to the secondary Kv·T times the transformer ratio.

Current TIF is measured by placing a current TIF coupling unit in the power circuit, as shown schematically in Fig. 3(B). It may be placed directly in a low-voltage circuit carrying 5 amperes or less, but it is usually placed in a current transformer secondary circuit. Its impedance at 60 cps is about $0.1+j0.15$ ohm so that it does not imposed a serious burden upon the current transformer. The conversion factor is computed in a manner analogous to that described above for the voltage coupler, assuming one ampere at 1,000 cps. The set reading is 20.2 for $I{\cdot}T=1$, and the chart in Fig. 4(B) is plotted accordingly.

The TIF of the current in the coupler is the I·T divided by the rms value of that current, and the TIF of the primary current is the same. The primary I·T is equal to the secondary I·T times the current transformer ratio.

Residual component voltage TIF of a 3-phase machine is measured by connecting the windings in delta with one corner open and measuring across the open corner. One third of the Kv·T product so measured, divided by the phase-to-neutral rms voltage in kilovolts, gives the residual component voltage TIF.

Residual current TIF is measured by paralleling the secondaries of the current transformers on the three phases, thus giving the vector sum of the three currents.

Table III. Harmonic Currents in Supply to 24-Phase Rectifier

Frequency	Line Current, Amperes	I·T	
		1935	1960
60	604.00	604	302
300	2.42	500	540
420	1.10	650	720
660	0.53	1,200	1,200
780	0.33	1,300	1,100
1,020	0.050	580	260
1,140	0.216	2,400	1,220
1,380	4.10	22,400	26,100
1,500	3.19	14,000	21,300
1,740	0.17	620	1,240
1,860	0.071	250	560
2,100	0.137	480	1,210
2,220	0.148	520	1,380
2,460	0.044	160	450
2,580	0.091	340	960
2,820	0.945	3,690	9,650
2,940	0.76	3,000	7,500
3,180	0.063	240	550
3,300	0.080	290	650
3,540	0.138	410	930
3,660	0.058	150	360
3,900	0.029	60	130
4,020	0.050	90	190
4,260	0.41	560	1,100
4,380	0.37	460	810
Root sum square		27,100	36,100

Comparison of Actual TIFs

Some comparisons of TIF computed from analyses of power system voltages and currents using the 1935 and 1960 TIF weightings are shown in the tables. Table II compares the open-circuit voltage TIF of several synchronous generators of various sizes and current design. It will be noted that use of the 1960 weighting increases the TIF in some cases and reduces it in others, depending on the particular distribution of the various harmonics involved.

Table III compares the I·T of a 13,200-kw 24-phase rectifier as measured

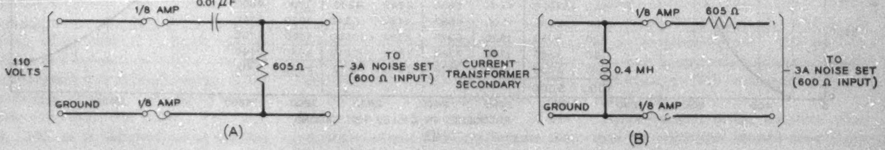

1/8 AMP
0.01 μF
110 VOLTS
605 Ω
GROUND
1/8 AMP
TO 3A NOISE SET (600 Ω INPUT)
(A)
TO CURRENT TRANSFORMER SECONDARY
1/8 AMP
605 Ω
0.4 MH
GROUND
1/8 AMP
TO 3A NOISE SET (600 Ω INPUT)
(B)

Fig. 3. TIF coupling units. A—Voltage. B—Current

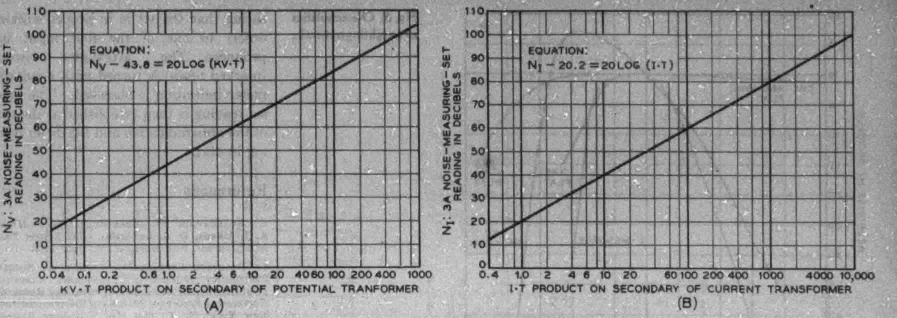


Fig. 4. Conversion charts of 3A set. A—Reading to Kv·T. B—Reading to I·T

in the 13.8-kv supply line. Table IV compares the I·T of one of the eight 6-phase elements comprising the same 24-phase rectifier. As might be expected, use of the 1960 weighting increases the I·T of the entire 24-phase rectifier because the 1960 weighting is greater in the region of the 23rd and 25th harmonics, which predominate in a 24-phase rectifier. Conversely, use of the 1960 weighting decreases the I·T of the 6-phase unit because the 1960 weighting is smaller in the region of the lower-order harmonics, which predominate in a 6-phase rectifier.

The 3A Noise-Measuring Set

The Western Electric *3A* noise measuring set is the only set at present available for measuring the 1960 TIF values, but other manufacturers may adopt the new weighting in the future.

Table IV. Harmonic Currents in Supply to 6-Phase Element of 24-Phase Rectifier

Frequency	Line Current, Amperes	I·T 1935	I·T 1960
60	74.	74	37
300	16.7	3,420	3,760
420	6.47	3,820	4,210
660	4.59	10,300	10,400
780	2.98	12,200	10,000
1,020	1.56	18,100	7,960
1,140	1.17	13,000	6,590
1,380	0.49	2,700	3,100
1,500	0.34	1,500	2,300
1,740	0.46	1,700	3,400
1,860	0.31	1,100	2,400
2,100	0.44	1,500	3,900
2,220	0.325	1,150	3,030
2,460	0.240	880	2,480
2,580	0.193	720	2,050
2,820	0.120	470	1,230
2,940	0.080	310	790
3,180	0.136	520	1,190
3,300	0.080	290	650
3,540	0.138	410	930
3,660	0.103	260	630
4,260	0.065	90	180
4,380	0.032	50	70
Root sum square		28,200	20,500

Since the set was designed to be useful as a general noise-measuring instrument, a brief description of some of its features may be of interest.

The basic requirement is that it give approximately the same reading on any two different noises which are judged to be equally interfering when heard over a 500-type telephone set. This is accomplished by means of its *C*-message weighting network, its detector (approximately square law), and dynamic response.

The appearance of the *3A* set is shown in Fig. 5. Its dimensions are approximately 8 by 11 by 7 inches and its weight about 15 pounds. This represents a considerable reduction in size and weight over previous noise-measuring sets; it was accomplished through the use of transistors and other miniature components.

An important physical innovation in the *3A* set is the plug-in arrangement for weighting networks. Two networks are packaged in a rectangular unit which may be inserted into a well in the panel in either of two ways, so as to use either network at will. The usual unit for general noise purposes and TIF contains *C*-message weighting and 3-kc flat weighting. An optional unit has 5-kc program weighting and 15-kc flat weighting. The meter has the same response at 1,000 cps with all four networks, and the set may therefore be calibrated with any network in position.

The *3A* set is powered by a small 45-volt battery. The battery life is expected to be 100 hours with 2 hours of steady operation per day.

Two separate single-tip and ring jacks afford access respectively to the monitor circuit which is the a-c output, and to the d-c output. When the a-c output is operated into 5,000 ohms or a higher impedance, no voltage clipping will take place until the peak output is 15 db higher than that corresponding to full-scale meter reading. The a-c output may be monitored with the special monitoring receiver furnished with the set. The noise being monitored may be heard at clearly audible volumes in the receiver for all meter readings from full scale down to zero db.

The d-c output may be used for oper-

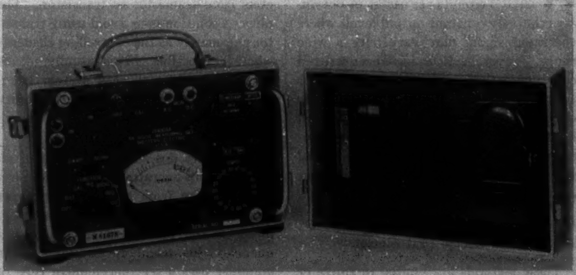


Fig. 5. The 3A noise-measuring set

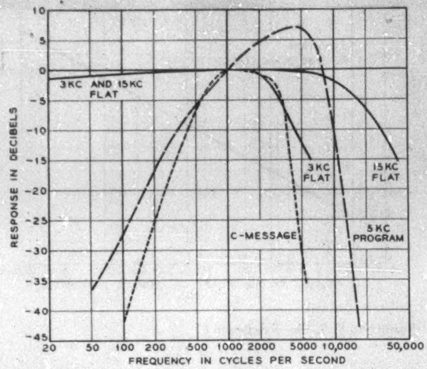


Fig. 6. Characteristics of weighting networks

ating a recording meter. The d-c voltage does not suffer compression up to a value of 1.4 volts across a load impedance of 1,100 ohms.

The sensitivity of the new set is −90 db below one milliwatt (10^{-12} watts or 24.5 μvolts across 600 ohms), and 90 db of amplification is provided. An internal oscillator gives one milliwatt at 1,000 cps for convenient calibration. The frequency characteristics of the weighting networks are shown in Fig. 6. The accuracy of response of the networks varies over the frequency range from ±0.1 db at 1,000 cps to as much as ±2 db on the steeper parts of the curves. The average on *C*-message weighting may be taken as about ±0.5 db, and the over-all accuracy of the *3A* set is about ±1 db.

Operation

The input of the set (see Fig. 5) may be connected to a telephone circuit or to a TIF coupling unit by means of a telephone plug, a banana plug, or an open-end pair of leads. The desired weighting network and the monitoring receiver are plugged in. The function switch is then turned through the BAT and CAL positions (assuring that the set is in proper working order) to one of the four input impedances. The attenuator switch (marked DBRN) is turned until a suitable meter indication is observed. The noise set reading is then established as the sum of the dbrn indication and the meter reading in dbrn.

References

1. EVALUATION OF MESSAGE CIRCUIT NOISE, **A. J. Aikens, D. A. Lewinski.** *Bell System Technical Journal*, New York, N. Y., July 1960.

2. RELATIVE INTERFERING EFFECTS OF SINGLE-FREQUENCY NOISES IN TELEPHONE CIRCUITS, **Joint Subcommittee on Development and Research.** *Engineering Report no. 32*, Edison Electric Institute, New York, N. Y., and Bell Telephone System, New York, N. Y., June 7, 1935.

3. THE TELEPHONE INFLUENCE FACTOR OF SUPPLY SYSTEM VOLTAGES AND CURRENTS, **Joint Subcommittee on Development and Research.** *Engineering Report no. 33*, Edison Electric Institute, and Bell Telephone System, Oct. 24, 1935.

4. SYNCHRONOUS GENERATORS, SYNCHRONOUS MOTORS, AND SYNCHRONOUS MACHINES IN GENERAL. *ASA C50.1-1955*, American Standards Association, Inc., New York, N. Y., 1955.

5. FREQUENCY WEIGHTING FOR MESSAGE CIRCUIT NOISE, **Joint Subcommittee on Development and Research.** *Engineering Report no. 45*, Edison Electric Institute, and Bell Telephone System. July 24, 1941.

6. A NEW MEASURING SET FOR MESSAGE CIRCUIT NOISE, **W. T. Cochran, D. A. Lewinski.** *Bell System Technical Journal*, July 1960.

7. SUSCEPTIVENESS OF SUBSCRIBER STATION SETS. **Joint Subcommittee on Development and Research.** *Engineering Report no. 46*, Edison Electric Institute, and Bell Telephone System, July 24, 1941.

8. MEASUREMENT OF TELEPHONE NOISE AND POWER WAVE SHAPE, **J. M. Barstow, P. W. Blye, H. E. Kent.** *AIEE Transactions (Electrical Engineering)*, vol. 54, Dec. 1935. pp. 1307–15.

Vol.
Pt. 1
196

A Transistor Automatic Gain Control System for Audio Frequencies

J. SHIRMAN
ASSOCIATE MEMBER AIEE

AUTOMATIC GAIN CONTROL (AGC) circuits are useful in obtaining relatively constant output levels where the input signal may vary over a wide range. Several articles have been written covering methods of controlling transistor gain in intermediate radio-frequency circuits.[1,2] The usual method employed is to change one of the transistor parameters by the application of an external signal. The two basic methods of gain control indicated are:

1. Emitter current (I_e) control.
2. Collector voltage (V_c) control.

These methods are limited in range of control and stability of output level because of inherent differences in parameters from transistor to transistor. The obvious solution is to avoid using transistor parameters as the controlling elements wherever possible. This usually means that a more complicated system is required. However, when good stability and wide operating range are required, no other solution exists.

This paper describes the design of an AGC system operating in the audio-frequency range from 300 cps (cycles per second) to 3,500 cps. This will be recognized as the pass band for audio frequencies in telephone systems. The amplifier uses a servomechanism type of closed loop feedback network to give a stable output level. One of the main advantages of this system is its ability to compensate internally for transistor parameter variations and supply voltage fluctuations in the amplifier section.

Circuit Description

This particular system was designed to compensate for input variations from −30 dbm (decibels referred to 1 milliwatt) to 0 dbm while maintaining the output at about +1 dbm. This system is used between an electronic field telephone and its associated mobile electronic dial switching system. All signaling is done with multifrequency tones

---

Paper 60-813, recommended by the AIEE Wire Communications Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting, Atlantic City, N. J., June 19–24, 1960. Manuscript submitted March 17, 1960; made available for printing August 12, 1960.

J. SHIRMAN is with the Stromberg-Carlson Company, Rochester, N. Y.

The author acknowledges the many contributions made by the engineers of the Telecommunications Department, Stromberg-Carlson Company; in particular, S. B. Weiner, A. F. Perkins, and J. H. Guenther.

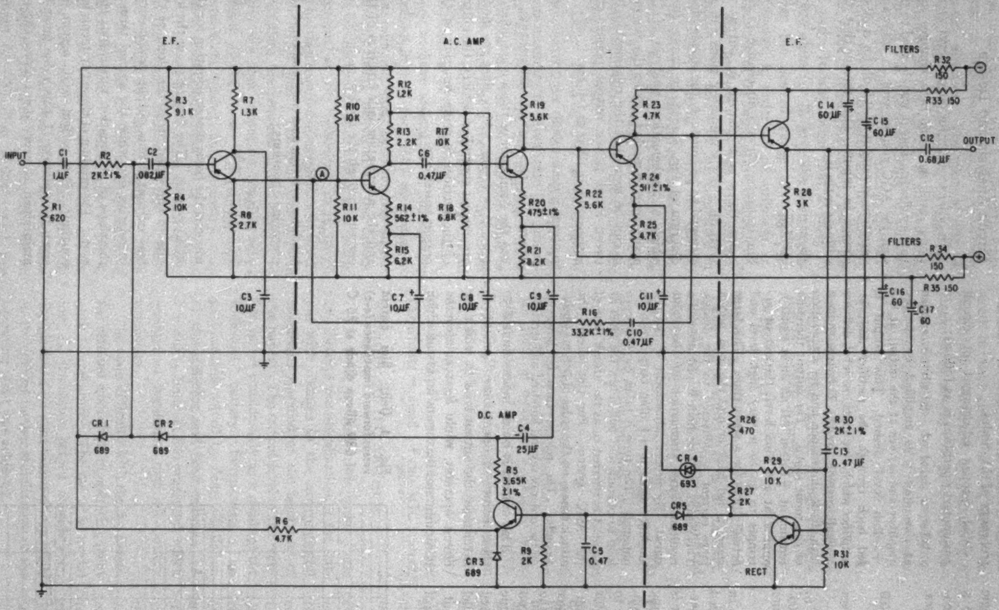

E.F.
A.C. AMP
E.F.
FILTERS
R32 150
R33 150
INPUT
C1 1µF
R1 620
R2 2K±1%
C2 .082µF
R3 9.1K
R4 10K
R7 1.3K
R8 2.7K
C3 10µF
R10 10K
R11 10K
A
R12 1.2K
R13 2.2K
C6 0.47µF
R14 562±1%
R15 6.2K
R17 10K
R18 6.8K
C7 10µF
C8 10µF
R19 5.6K
R20 475±1%
R21 8.2K
C9 10µF
R16 33.2K±1%
C10 0.47µF
R22 5.6K
R23 4.7K
R24 511±1%
R25 4.7K
C11 10µF
C14 60µF
C15 60µF
R28 3K
C12 0.68µF
OUTPUT
FILTERS
R34 150
R35 150
C16 60
C17 60
D.C. AMP
C4 25µF
CR1 689
CR2 689
R5 3.65K ±1%
R6 4.7K
CR3 689
R9 2K
C5 0.47
R26 470
CR4 693
R29 10K
R27 2K
CR5 689
R30 2K±1%
C13 0.47µF
R31 10K
RECT

Fig. 1. Complete AGC system schematic

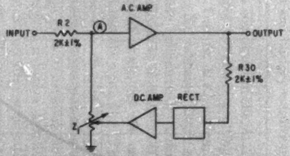

INPUT
R2 2K±1%
A
A.C. AMP
OUTPUT
R30 2K±1%
DC AMP
RECT
Z1

Fig. 2. Simplified block diagram representation of AGC system

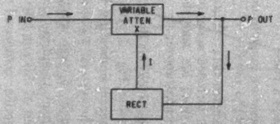

P IN
VARIABLE ATTEN. X
P OUT
I
RECT

Fig. 3. Basic AGC circuit showing relationship of main parameters

originating in the telephone. These frequencies all fall within the audio spectrum. Signaling tones and voice could be received at the switching center at any level from 0 dbm to −30 dbm depending on the condition of the telephone and the connecting line length (up to 8 miles). Operation of subsequent logic functions in the switchboard require a +1.5 dbm level ±1.5 db over wide environmental and component ranges. These are as follows:

1. Battery voltage variable from ±10.0 volts to ±14.5 volts.
2. Temperature range from 30 F (degrees Fahrenheit) to 130 F.
3. A predicted beta variation from 20 to 200 which included aging over a 200,000-hour (20-year) life and the temperature range indicated above.

A complete circuit diagram is given in Fig. 1, while Fig. 2 gives a simplified block diagram indicating the main sections of the system. The a-c amplifier is a fixed gain 3-stage circuit containing loop feedback to aid in gain stabilization. The output of this amplifier is fed back to a peak rectifier comprised of a transistor and diode. This voltage is then converted to a current in the d-c amplifier stage which supplies the current to the variable impedance Z_1. Two silicon diodes are used with a wide dynamic impedance range in the forward biased direction. The two diodes are in parallel when viewed in an a-c equivalent circuit. The use of two diodes extends the range of operation and helps stabilize against diode variations from unit to unit. The input signal is attenuated by the variable L pad comprising the fixed resistor R_2 and the variable Z_1. The operation of this pad is such that the voltage at point A in Fig. 2 is always constant over the control range of the diodes.

With proper adjustment of circuit parameters operation is as follows:

When a signal at −30 dbm is applied at the input it is amplified by the a-c amplifier and arrives at the output at the +1-dbm level desired. This output is fed back through the rectifier and d-c amplifier which, due to its low level, delivers a small current to Z_1. This results in a high diode impedance and no attenuation of the input signal. The output therefore remains at +1 dbm. If an input signal is applied at a 0 dbm level, the output attempts to go to a very high level and actually does for a short time. This large signal is fed back around the loop, and a large current is delivered to the diodes. This results in a low diode impedance and a resultant large attenuation of the input signal. By proper adjustment of the d-c amplifier again a +1 dbm output can be obtained. The time it takes for the servo loop to operate and pull down the level at point A is referred to as the attack time of the AGC system. For this circuit the attack time is on the order of 5–8 ms (milliseconds). The release time is, however, on the order of several hundred milliseconds.

This system is by no means limited to these points of operation. Resistor R_2 determines the output level at which the circuit operates, while R_{30} regulates the minimum and maximum inputs at which regulation is obtained. By judicious adjustment of R_{30}, R_2, and the gain of the d-c amplifier, the system can be made very flexible with regard to output level and range of input level which can be controlled.

Circuit Analysis and Design

In order to allow for output level variations resulting from beta changes and battery fluctuations it was decided to design the AGC system with an allowable variation in output of 0.5 db for a 30-db change in input. This means that the stiffness ratio is:

$$\frac{dP_o}{dP_i}=\frac{1}{60} \tag{1}$$

For instance, for a 1-db change in the input there is a change in output of 1/60 db. If the amplifier gains can be assumed to be constant, the simple block diagram of Fig. 3 may be drawn from which a mathematical analysis of the system can be made.[3] From the diagram it can be seen that the relationship between input power, output power, and attenuation in db is as follows:

$$P_o=P_i-X \tag{2}$$

where X is the variable attenuation, x, less the fixed gain of the a-c amplifier, k (i.e., $X = x - k$). Also, since the diodes are current controlled devices, the output of the rectifier is some function of the output power level:

$$I=f_1(P_o) \tag{3}$$

and the attenuation is a function of the current through the diodes:

$$X=f_2(I) \tag{4}$$

By differentiation and combination of the preceding equations:

$$\left(\frac{\partial I}{\partial P_o}\right)\left(\frac{\partial X}{\partial I}\right)=59 \tag{5}$$

which is the basic design equation for the AGC system. The first term on the left side of the equation is the slope of the bias current versus output power curve in relative db, while the second term is the slope of attenuation versus bias current. The product of these two terms is a constant. For perfect AGC action (as given by equation 1), this value must be maintained at every point along the re-

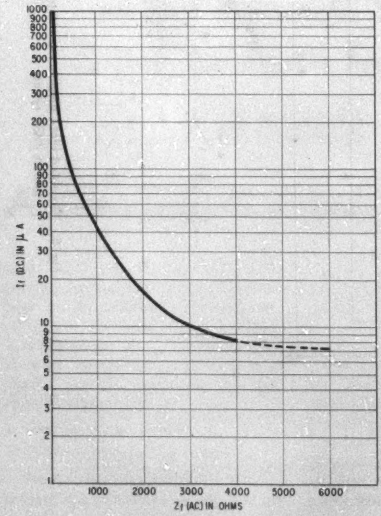


Fig. 4 (left). Bias current versus forward impedance (a-c) of 689 silicon diode at 25 C

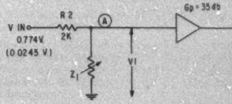


Fig. 5 (right). Equivalent circuit for calculating diode impedance

quired operating range. As will be shown later, experimental results are very close to the desired ideal values.

Diode Operation

The first step in the design procedure was to find a diode with a sufficiently wide dynamic impedance range in the forward biased direction. The plot of Fig. 4 shows bias current versus forward impedance of the silicon diode selected. Due to the interdependence of various sections of this system, considerable empirical work was necessary in selecting the value of fixed resistance to use in the *L* pad. It was found that a value of 2,000 ohms for R_2 gave the best results over the desired range of operation. However, in order to get sufficient attenuation of the input at the 0-dbm level, it was necessary to use two diodes. Diodes *CR1* and *CR2* are connected in a-c parallel but in a series d-c circuit. This gives the benefit of a lower a-c impedance at a given current level. At the same time, by feeding direct current in series, differences between the two diodes are balanced out.

Before deciding over what range of bias current the diodes must operate, it was necessary to investigate the gain requirements of the a-c amplifier. It is obvious that the minimum gain must be 30 db since that is the lowest input level requiring control. In order to get rapid AGC action at the low end, it was necessary to have several db of attenuation present in the *L* pad. This meant that at the minimum input level some current was already being delivered to the diodes. It was decided to have 4 db of attenuation present at the −30-dbm input level. This then required an amplifier with about 35 db of gain. The next step was to calculate the required *L* pad needed to give +1 dbm output for a 0-dbm input signal (0.774 volt). Fig. 5 gives the equivalent circuit and necessary data used in the calculations.

It is seen that V_1 must be 0.0155 volt at this input. This means that Z_1 must be about 40 ohms. Therefore, each diode should present an impedance of about 80 ohms. Looking again at Fig. 4, it was seen that the required bias current for this value of forward impedance is 750 μa (microamperes). In the same manner the value of Z_1 required for an input of −30 dbm (0.0245 volt) was calculated. In this case the output level could drop to +0.5 dbm (0.825 volt). This gave a value for Z_1 of about 3,000 ohms or an individual diode impedance of 6,000 ohms and bias current of 7.5 μa. These two bias currents gave the range of currents required from the d-c amplifier. This amplifier is strictly a linear device, that is, the output current is a linear function of the input. In order to get rapid AGC action it was necessary to have a nonlinear resistance as the variable element.[4] The silicon diodes used in this circuit act as linear to logarithmic converters by converting a linear input current change to a logarithmic variation in dynamic impedance. The ideal relationship for the conversion is given by:

$$A=\epsilon^{CB} \tag{6}$$

where *A* and *B* are variables and *C* is a constant. The current voltage relationship for a diode is given by:

$$I=I_s(\epsilon^{qv/kt}-1) \tag{7}$$

neglecting the *IR* drop of the diode. At room temperature q/kt is a constant. Also, at the normal voltages encountered across a diode:

$$\epsilon^{qv/kt}\gg 1 \tag{8}$$

which means that equation 7 reduces to

$$I=I_s\epsilon^{Cv} \tag{9}$$

where I_s (the saturation current) and *C* are constants. This equation is now in the form of equation 6. Fig. 6 is the well-known *V-I* diode curve for the *689*-type used over the range where equation 9 holds. The linear to logarithmic conversion is seen to be about 0.1 volt/decade and to be constant over the range of currents required for this AGC system.

Rectifier and D-C Amplifier

The main problem in designing these two circuits was stabilization of the output with changes in the environmental conditions specified at the beginning of the paper. The output level of the rectifier varied considerably as a result of the indicated battery fluctuations. This problem was solved by Zener regulation of the collector and base voltages. A 7.5 ±5% Zener (*CR4*) was used which guaranteed regulation over all battery conditions.

The main difficulty with the d-c amplifier resulted from shifts in operating point with temperature. The use of a special low I_{co} transistor helped considerably. By placing a forward biased silicon diode in the emitter further stabilization of the operating point was achieved.

A-C Amplifier

Once the gain requirements of the a-c amplifier were determined the design procedure was quite straightforward. The amplifier consists of three stages, each

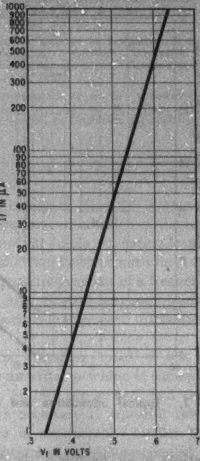


Fig. 6. Forward voltage versus forward current curve for 689 silicon diode at 25 C

containing emitter degeneration. All three stages are designed with d-c operating point stabilization through the use of base biasing resistors. Over-all loop feedback of about 10 db around the a-c amplifier is also included. Each stage has about the same gain, approximately 12 db. The amplifier alone exhibited a stability of better than ±0.4 db over all the beta and battery conditions indicated before.

When connected in the AGC system a very important and interesting result was noted. Due to AGC action, the variation in gain indicated above was almost completely eliminated. Changing the three transistors from betas of 20 to 150 (the highest available) the output level changed less than 0.1 db. This resulted from the operation of the AGC loop in maintaining the output at a fixed level. The high beta transistors acted in the same way as a higher amplitude signal resulting in greater attenuation in the AGC loop. The converse is true in the case of the low beta units. For this reason the AGC system is quite insensitive to beta variation, the only transistors having any effect being the two in the AGC feedback loop.

Experimental Results

In testing the system it was found that the variations in dynamic impedance of the diodes had serious effects on the operation of the a-c amplifier. Also

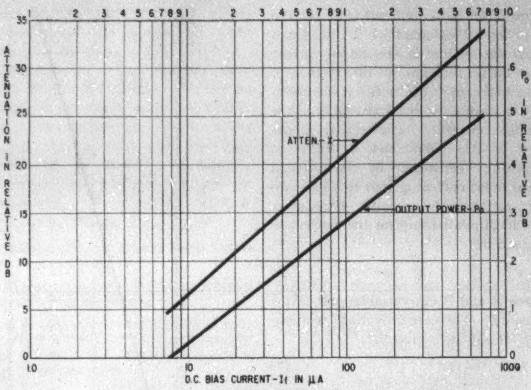


Fig. 7. Typical attenuation and output power versus bias current curves found during development

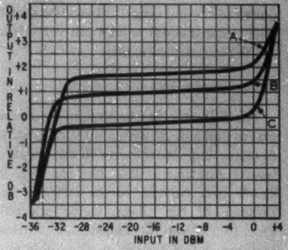


Fig. 8. Input versus output for complete AGC system

A—Volts: ±14.5; B maximum = 150; T_A = 135 F
B—Volts: ±12.3; B nominal = 60; T_A = 75 F
C—Volts: ±10; B minimum = 20; T_A = 20 F

variations in a-c loading of the AGC affected the operating range of the circuit. In order to eliminate these effects emitter follower circuits were included at the input and output of the a-c amplifier to isolate it from impedance variations at its terminals. The emitter follower output means that the system has a very low output impedance. However, due to loading effects, the minimum load resistance before serious distortion occurs is about 1,000 ohms. On the input side, the 620-ohm shunting resistor, R_1, has the effect of stabilizing the input impedance variation due to the variable *L* pad. For a 30-db input voltage change the input impedance varies from about 450 ohms to 580 ohms or a total of about 25%.

Fig. 7 shows typical curves of attenuation and output power versus diode bias current. It is the slope of these curves which must satisfy equation 5. A check will show that correlation is quite close over most of the operating range.

One factor which tends to limit the usefulness of the circuit is distortion. Due to the fact that the diodes are used over a wide dynamic range they tend to introduce distortion at the higher input levels. Distortion at −30 dbm input is about 5%. However, at an input of 0 dbm the total rms distortion increases to a value of 10% to 15%. No simple solution to this problem has been found. It is believed that a bridge diode network instead of the parallel connection would be a considerable improvement.

Fig. 8 is a set of curves of input versus output for the complete AGC system. The *A* and *C* curves are "worst condition" results and show that the design requirements were easily met. It was found that variations in the output level occurring with differences in beta and temperature were almost exclusively the result of effects of those conditions on the rectifier and d-c amplifier circuits. It was also noted that if the minimum beta was increased from 20 to 60 (the nominal value indicated), curve *C* is almost equal to curve *B*. This shows that the greatest variation in output level is due to the spread in transistor beta, particularly in the feedback loop. By raising the minimum beta requirements it is possible to design this system with a total output variation of less than 1.0 db. It should also be noted that temperature has some effect on the range in which the diodes act as good logarithmic converters. This can be noted in Fig. 8 at the lower end of the operating range. The operating range at 20 F is about 2 db more than at 135 F.

The frequency response of the system is flat, ±0.5 db from 400 cycles to 10 kc. Fig. 9 shows a nominal frequency response curve. The curve shows that the system displays a 3-db peak at about 150 cycles. This peak results from some of the *R-C* time constants in the system. Since this system was intended for use primarily over the telephone audio band of 300 to 3,500 cps, this was not considered a problem.

Some low-frequency oscillations were encountered along with feedback of pulses through the battery supply from other circuits. To eliminate these problems, four decoupling circuits were used as indicated in Fig. 1. These solved all the problems of intercircuit interference as well as preventing any oscillations due to mutual feedback between various parts of the AGC system.

Conclusions

Although this paper outlines the design of a particular AGC system it is evident that the procedure and block diagram configuration are applicable to a wide range of design requirements. The principal problem centers on finding the proper diodes to do the job. This is not always easy, but considerable help can usually be obtained from diode

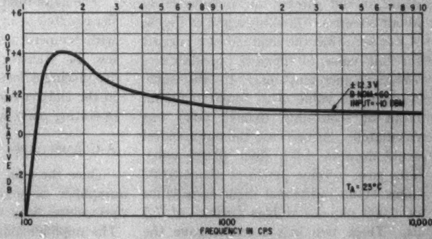


Fig. 9. Frequency response of agc system

manufacturers. Wider ranges of operation should be possible by using diodes in series or bridge arrangements.

References

1. Automatic Gain Control of Junction Transistor Amplifiers, F. H. Blecher. *Proceedings*, National Electronics Conference, Chicago, Ill., vol. 9, Sept. 1953, pp. 731–37.

2. Automatic Gain Control of Transistor Amplifiers, W. F. Chow, A. P. Stern. *Proceedings*, Institute of Radio Engineers, New York, N. Y., Sept. 1955, pp. 1119–27.

3. A Transistorized Compandor, J. C. Perkins, Jr., D. A. Perreault, A. F. Perkins. *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 77, 1958 (Jan. 1959 section), pp. 791–97.

4. Transistor Electronics (book), R. B. Hurley. John Wiley & Sons, Inc., New York, N. Y., 1958.

5. Silicon Diodes as Logarithmic Elements, J. Golahny. *Raytheon Technical Bulletin*, Raytheon Manufacturing Company, Waltham, Mass., Oct. 1957.

6. Adjustment Speed of Automatic Volume Control Systems, A. W. Nolle. *Proceedings*, Institute of Radio Engineers, July 1948, pp. 911–16.

Datacom Unit—An Experimental Data Transmission Subset

R. E. STOFFELS
ASSOCIATE MEMBER AIEE

THE NUMBER of miles of telephone wire in the United States today runs well over the 240 million mark. This is enough to run 1,000 wires between here and the moon, or, if extended as a single pair, enough to carry on a telephone conversation with the sun. Also there are in use today, in this country alone, well over 60 million telephone instruments.

The almost unbelievable magnitude of these figures suggests that future growth in the telephone industry will be mainly horizontal rather than vertical. That is, the growth will be more in usage than in number of miles of wire, or number of telephone instruments. One field which promises to provide such expansion is that of data transmission. Data that were formerly collected with paper and pencil, acted upon by trained personnel, and transmitted by mail or by hand, are now being collected and acted upon automatically, and, as a matter of course, must be transmitted automatically.

Transmission

The transmission of digital data over telephone lines is not new; in fact, the world's vast teletypewriter networks use, for the most part, existing telephone lines. In order to make this transmission possible, however, a language translation is necessary. This is because of the basic characteristics of a telephone line, or, to be more precise, a telephone channel.

A telephone channel is not capable of transmitting all frequencies from 0 cps (cycles per second) up; if it could it would not be economical, for a telephone channel was designed to carry a human voice, and in order to do this only those frequencies from approximately 250 cps to 3,000 cps have to be reliably transmitted. Consequently, a telephone channel is essentially a band-pass filter which severely attenuates those frequencies below approximately 250 cps and above 3,000 cps. (These cutoff frequencies will vary from circuit to circuit.) Because of the low-frequency limitation of the telephone channel all pulses emanating from a data transmitting machine must be in the form of alternating current of a frequency over 250 cps. Several language translations, commercially employed, will be mentioned later.

The operating speed of the popular teletypewriter machine is somewhat less than 80 bits per second, and the bandwidth necessary to transmit this intelligence is considerably smaller than the bandwidth of a telephone channel. Other business machines, which transmit data at higher speeds, require proportionately larger bandwidths. The maximum speed attainable, however, is limited by such factors as impulse noise, frequency band shift, and delay distortion.

There are two types of data transmission systems in common usage today; the synchronous system and the start-stop system.

In the synchronous system the re-

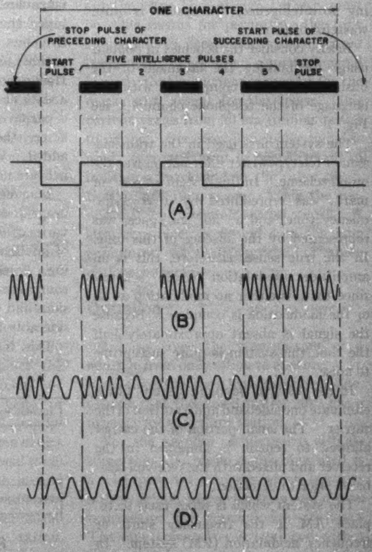


Fig. 1. Examples of transmission schemes

A—Mark space
B—AM
C—FM
D—Phase modulation

Paper 60-790, recommended by the AIEE Data Communication Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting, Atlantic City, N. J., June 19–24, 1960, and represented for discussion only at the Fall General Meeting, Chicago, Ill., October 9–14, 1960. Manuscript submitted April 5, 1960; made available for printing August 15, 1960.

R. E. Stoffels is with Automatic Electric Laboratories, Inc., Northlake, Ill.

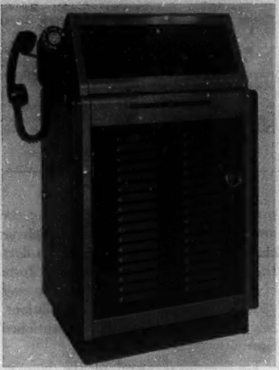


Fig. 2. Datacom unit

ceiving business machine is, in some manner, made to run at exactly the same speed as the transmitting business machine; often two stable oscillators are used.

In the start-stop system the receiver is allowed to run at a speed slightly faster than that of the transmitter. It comes to a complete stop every so often (usually once each character) and waits for the next signal from the distant transmitter. Because the transmitted signal must contain a START and a STOP pulse carrying no intelligence, this system is quite wasteful of time.

Briefly consider the schemes in popular usage for making the language of the business machine compatible with the language of the telephone channel;[1] see Fig. 1.

The system first used in the transmission of teletypewriter information was the on-off scheme. In this system a "yes," or mark, was represented by a voice-frequency tone, and a "no," or space, was represented by the absence of this tone. In the true sense, therefore, this is an amplitude modulation (AM) scheme, since there is either no modulation at all or the modulation is complete. Because the signal is absent approximately half the time this system is quite susceptible to noise.

In the vestigial sideband system, filters eliminate one sideband and a portion of the carrier. The small portion of the carrier allowed to remain is amplified in the receiver and mixed with the received sideband prior to detection.

The system which is often used to replace AM is the frequency shift, or frequency modulation (FM) system. In this scheme a "yes," or mark, is represented by one frequency and a "no," or space, is represented by a second frequency. The technical considerations of this system are very similar to that of the AM system, except that the FM scheme is considerably less vulnerable to noise.

The latest modulation scheme, and one which appears to have a great future, is phase modulation. In this scheme a continuous carrier wave is transmitted, but its phase is shifted depending upon whether the intelligence bit is a mark or a space.

Terminal Devices

The devices which will take intelligence and transmit it over a telephone channel and receive this intelligence at the distant station have been mentioned and also the speed capabilities of the telephone channel, the problems encountered in transmitting intelligence, and the modulation schemes which are commonly used. A brief investigation follows of the type of machines which are used to generate this intelligence at one end and to present it, at the other end, in a usable manner.

There are in use today thousands upon thousands of teletypewriter machines. These machines normally use a start-stop code, with the stop pulse either exactly as long as an intelligence pulse, or, more commonly, 1.42 times as long. These machines usually use five-channel paper tape; that is, each character consists of five bits. In the binary system this permits the use of 32 characters ($2^5=32$). If more possible combinations are required a sixth channel is added. If, in addition, a parity check is required of each character, then an additional channel is added. Very often, in fact, eight channels are used in transmitting data.

Magnetic tape is coming into wide usage today, and the type of tape used varies considerably, both as to the number of intelligence channels and the bit spacing. Punched cards are used by several companies, and both the physical dimensions and the number of holes varies considerably from one to the other.

This is mentioned to point out that the companies in the data field, the codes used for transmitting data, and the means used for recording this data vary enormously. It becomes quite apparent that any device used by the telephone company to transmit data must not try to replace these business machines, but rather must accept data from them at one end and present data to them at the other end. It must perform a translation from the language of the business machine to a language capable of being transmitted over telephone channels and back again to the language of the business machine.

Datacom Unit

In designing a data transmission device it was recognized that no one unit would work with all business machines; for instance, simple circuits adequate for slow-speed operation would be completely unsatisfactory for use with high-speed machines.

Because of the large number of teletypewriter machines in general use it was this machine which was considered in our design program. It was felt that as many special features as possible should be included with the device, even though in most applications many would not be useful.

The result of this design effort was the "Datacom" digital data communication set, a unit which we feel is a big step forward in meeting the requirements for data transmission; see Fig. 2.

The Datacom unit is a transmission device; it accepts information from a business machine and does not try to replace the machine; see Fig. 3.

In the Datacom unit an FM modulation system is used; equipment utilizing FM modulation is commercially available and has been perfected by manufacturers to a high degree.

Telephone lines are used on a demand basis; thus leased lines associated with the equipment on a permanent basis are unnecessary. Data are accepted from the associated business machine in a sequential form; therefore, the number of intelligence channels used by that business machine is immaterial. The sys-

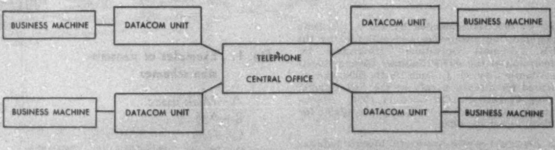


Fig. 3. Block diagram showing Datacom units in relation to central office

Vol.
Pt. 1
196

tem is nonsynchronous and the expense of local controlled oscillators is eliminated.

The Datacom unit consists of: tone equipment and equipment required to convert the data communication language of the business machine into a "tone" language compatible with the voice-frequency telephone lines; control equipment required for the necessary dialing, automatic answering, switching from voice communication to data communication, and resetting of equipment; and loop equipment which serves to connect the particular business machine to the Datacom unit.

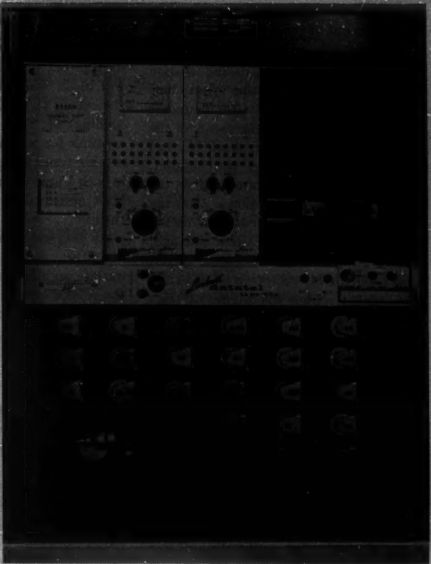


Fig. 4. Inside view of Datacom unit showing tone and control equipment

TONE EQUIPMENT, FIG. 4

The tone equipment used is standard Lenkurt type *23A* telegraph carrier equipment[2] and is entirely transistorized, operating on a FM basis. The carrier frequency used is 1,955 cps, and the mark and space frequencies are formed by shifting the carrier frequency between values which are 35 cycles above and below the nominal midband frequency. Consequently the midband frequency of 1,955 cps never actually appears. Both levels may range from +6 dbm (decibels referred to 1 milliwatt) to −40 dbm. The equipment is designed for a line impedance of 600 ohms.

TRANSMITTING EQUIPMENT

In the transmitting equipment, a diode modulator is used to insert or remove a shift capacitor from the oscillator tank circuit, thus minimizing the effect of the line to the business machine. The oscillator can best be described as a regenerative flip-flop whose oscillations are limited by clipping in both transistors. This stage feeds a buffer amplifier; the output is fed through a 2-transistor amplifier, and an output transformer feeds the transmission line.

RECEIVING EQUIPMENT

In the receiving equipment, the signal is fed through a pad, if required, then through a band-pass filter and a single stage of amplification. It is then severely clipped and fed through a transformer to a push-pull stage. The output of this stage feeds a discriminator of the double tuned series-resonant type. One mesh is tuned to a frequency close to the mark frequency, and the other mesh is tuned to a frequency close to the space frequency. The output of each mesh is bridge-rectified and two currents are combined differentially. The resultant signal then feeds two power transistors connected as d-c amplifiers, and this final output drives a polar relay. The polar relay is used to operate the receiving portion of the business machine.

CONTROL EQUIPMENT, FIG. 4

The function of the control equipment begins when a call is initiated and ends when the line is released. It includes the initiation of a call from a standard telephone dial, automatic answering of the distant station in response to ringing current, and control of the tone equipment. The equipment consists of a control panel from which all operations are performed, control relays, transistor circuits used for timing purposes, and a telephone with means for automatically cutting the handset transmitter in and out of the circuit at the proper time. Thus it is possible at all times to monitor the voice-frequency signals being transmitted and received and, in certain instances, to carry on voice communication.

Because this system provides unattended service, certain additional equipment had to be added to replace the spoken "hello" of a normal telephone conversation with an electronic "hello" or "go ahead." This additional equipment consists of a unit of FM tone equipment, thus providing two additional tones for supervisory functions. The second of these tones is used to indicate an alarm condition, and will sound a buzzer to call the operator at the distant station. Voice communication can then be carried on.

LOOP EQUIPMENT

Since all business machines do not put out the same kind of signal, there must be some means for permitting the tone equipment to accept the specific signal being used. Very often a neutral signal is used; that is, current is present for the mark, and absent for the space. This current varies from 20 ma (milliamperes) to 60 ma. Occasionally a polar signal is used; current is sent in one direction to indicate a mark and in the other direction to indicate a space. Again the current can vary between several values. A strapping block is provided in the loop equipment to make several of these options possible.

Operation

Let us trace the transmission of a message from one station to another. To make this situation a little more complete, assume that the called station also has a message for the calling station.

For the purpose of discussion also assume that the business machine used is a teletypewriter machine which contains a page printer to receive data and a tape transmitter for transmission purposes. The tape transmitter is equipped with a "tape-out" pin, which indicates when the machine is out of tape, that is, a message has ended.

The operator at the calling station momentarily depresses the START button, thus applying power to the electronic equipment and starting the motor of the associated teletypewriter machine, and listens for dial tone on the handset. Upon hearing the dial tone he dials the number of the distant station. After the central office has selected this distant station, ringing current is sent out and, instead of operating a telephone ringer, it operates a relay receptive to this ringing current. The operation of this relay applies power to the tone equipment and turns on the motor of the teletypewriter machine.

When the called Datacom unit is ready for operation it sends a go-ahead tone. The Datacom unit at the calling station receives the tone and automatically starts the teletypewriter tape transmitter. The tape passing through the transmitter, with the aid of the loop equipment, modulates the FM tone equipment, thus transmitting data to the telephone channel. When the distant station receives this signal it reconverts it into data communication signals compatible with the teletypewriter machine; this causes the data to be typed as page copy. Simultaneously, the calling station also reads the signals from the line and causes a monitor copy to be printed by its page printer. (Notice that the only function performed by the operator was the initial dialing of the telephone connection. It is interesting to speculate on the possibility of making the calling station as well as the called station unattended; the simple addition of an automatic dialing device, in conjunction with a clock, is all that would be required.)

After transmission of the message, the tape-out pin rises on the tape transmitter. This stops the transmitter and initiates the transmission of a go-ahead signal to the called station. (The circuits are so arranged that the Datacom unit initiating the go-ahead signal cannot receive it.) The Datacom unit at the called station receives this signal and starts its teletypewriter transmitter. The message being sent from the called station is received at the calling station and, as before, at the called station for a monitor record.

When transmission of the message is completed, another go-ahead signal is initiated, this time by the called station. The calling station will receive this tone, but because its tape-out pin will be in the raised position it will transmit the go-ahead signal back to the called station. Since the called station has no message its tape-out pin will be up; therefore, it will again send the go-ahead signal to the calling station. This back-and-forth sending of the go-ahead signal will continue for several seconds. At the end of this time, signal-detecting equipment in each Datacom unit will "time out," and both units will release themselves from the line.

As previously mentioned, an operator at either station can, at any time, initiate an alarm condition by operating the ALARM button on the Datacom unit. This causes a different tone to be transmitted to the line, which is received by the equipment at both stations. This alarm signal halts transmission, prevents time-out from the signal-detecting equipment, lights an alarm lamp, sounds a buzzer, and connects the transmitters of the telephones. Operators at both stations can now converse in a normal manner. This alarm condition is cleared by an operator at either station by operating the GO-AHEAD button to allow further data transmission. Notice that action by operators at both stations is not required; the operator at one station can cause an alarm in order to attempt a conversation and if the other station is unattended this same operator can release the alarm. Should the alarm persist for one minute without being answered, the particular unresponding Datacom unit releases itself from the line.

Another operating feature of the Datacom unit is called "delayed alarm." At times it is advantageous to discuss a message with the distant operator before transmitting it. This is accomplished by operating the DELAYED ALARM key. In this case the receipt of the go-ahead signal will, instead of starting the transmitter, cause an alarm signal to be sent out.

It was recognized that different users would have many different communication requirements, and therefore the Datacom unit was designed to provide maximum flexibility of operation. Several of these possible modes of operation are discussed in the following.

KEYBOARD TRANSMISSION

It is quite possible that in certain applications transmission to the line directly from a teletypewriter keyboard would be desired. This has the advantage that the tape perforators and tape transmitter are not needed; however, it is wasteful of line time, since no operator can type as fast as a tape transmitter can transmit. Also it is not possible to proofread the message before transmission. At any rate, the Datacom unit accommodates keyboard transmission without any difficulty.

OPERATION IN A MANUAL TELEPHONE SYSTEM

This equipment can be operated with a manual telephone exchange. In such a case the attendant initiating the call operates the SUBSCRIBER DIAL-OPERATOR key to the Operator position, and then pushes the START button. This causes the transmitter of the telephone to be connected to the line; the attendant can now pass instructions to the operator at the central office. When the distant station answers, the handset is replaced, the GO-AHEAD button is pressed, and transmission begins automatically.

SECURITY SYSTEM

In certain applications of the fully automatic system positive identification of the distant station would be required before starting transmission. Such a security system has been added. In this case transmission of a message is contingent upon the receipt of a predetermined code from the distant station. (In a telephone system the recognition of the called party's voice usually provides the only security measure.)

AUTOMATIC DIALING

The delivery of collection of information during night-time hours would permit the user to take advantage of low rates, but the necessity of having even the calling operator on duty at this time might make such operations undesirable. The addition of an automatic dialing device, e.g., the Dialaphone, in conjunction with a timing device, would make such night-time operation completely automatic.

Conclusions

In the not-too-distant future the telephone industry will be as intimately interested in the transmission of data over their lines as they will be with the transmission of voice. More and more there will be a requirement for devices capable of translating the language of business machines into the language of the telephone channel. Speed and reliability requirements will be increased and, as a result, telephone channels will be upgraded and new electronic equipment will be designed. The exact direction of this design work is as yet unknown and will very likely remain so for a considerable time to come. Design, test, and redesign must determine the devices which will best serve the industry. The concept followed in the development of the Datacom unit is simply one solution; a very logical one we believe, but nevertheless only one.

Just as the amount of data being used by man is increasing by leaps and bounds, the facilities for the transmission of data must follow. The point-to-point transmission of these data, accurately and rapidly, is a challenge to the entire telephone industry.

References

1. An Introduction to Carrier Current Communications, J. N. Petrie, H. E. Clapham. Automatic Electric Company, Northlake, Ill., 1949.

2. Design Features of a New Frequency Shift Carrier Telegraph System, M. L. Stephens. *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 78, 1959 (Jan. 1960 section), pp. 927–34.

The New Nation-Wide Telephone Numbering Plan

O. MYERS
MEMBER AIEE

THE PRESENT NATION-WIDE telephone numbering plan, adopted in 1947, was expected to last beyond the year 2000, but now it appears that it will be outgrown by about 1975. The telephone industry in the United States and Canada has, therefore, adopted a new plan with several times the code capacity of the present one which should last well into the next century.

The new plan, shown in Table I, provides uniform dialing procedures for all types of central offices. Under the present plan, procedures vary with the type of office. The following are also provided:

1. A prefix for customer dialed person-to-person, collect, credit card, and other calls requiring operator assistance. The recommended prefix is "0" (zero).
2. A simplified toll office access code for DDD (direct distance dialed) calls in step-by-step central offices. With the present plan a step-by-step central office customer making a DDD call uses a 3-digit access code such as "112" to reach the toll office. The access code may be followed by as many as ten digits giving a maximum of thirteen digits. With the new plan, the recommended DDD access code is the single digit "1" and the maximum number of digits is therefore eleven.
3. A prefix to prevent error when local-intended calls reach toll points in cities using telephone systems like no. *5* and no. *1* crossbar which store and process the dialed digits (common control type). The new plan requires the use of a prefix, "0" or "1," on all toll calls. Hence a wrong office code digit on a 7-digit local call (no prefix being dialed) cannot convert a local-intended call into a toll call.

Background

The present nationwide numbering plan operates with 152 area codes of the form A1X or A0X (A represents 2 to 9; X represents 0 to 9). Of the potential 160 codes, eight are used for service codes in the series A-1-1. For example, the information code in a common control city is 411. The area code is generally followed by two letters of an office name and five numbers. See Table II for present and ultimate capacities.

With the present practice of using central office names, each area has a capacity of about 540 office codes. It would seem that there should be 640 possible office codes. (Eight dial pulls for the first digit times eight dial pulls for the second digit times ten for the third digit. Only eight holes on the dial have letters.) In practice, however, there are only about 60 usable letter combinations for the first two digits instead of 64. This is because of the difficulty in finding names to fit the dial pulls 5-5, 5-7, 9-5, and 9-7 (5 corresponds to JKL, 7 to PRS, and 9 to WXY). Also, since there is considerable confusion between the letter "O" and the numeral zero, the latter is usually avoided in the central office designation. This leaves us with 60 usable letter combinations multiplied by nine numerals for a total of 540 codes.

Each office code has capacity for 10,000 telephone numbers. The original plan of using office names was recently amended. All-number calling (ANC) is now recommended and 640 codes are available at present in areas that elect to use ANC. The combination of ANC and new dialing procedures will produce 792 office codes.

Eighty-six of the 152 area codes were assigned when operator nation-wide dialing was started in 1947. The remaining 66 area codes were expected to care for many years of growth. However, there has been a telephone explosion in the United States. In the decade 1950–1960, while the United States population grew from about 150 million to 180 million, telephones grew from about 41 million to 71 million. Growth during the next 15 years is expected to be even greater. Telephones are expected to double in number from 71 million to 142 million while population may increase to 237 million.

These data are somewhat misleading. Telephones rather than telephone numbers have been discussed while only the latter affect numbering capacity. Much telephone growth will merely result in an increased number of extensions rather than in more numbers. Some of the

Table I. The New Numbering Plan

	Ultimate Stage	Initial Stage
Home area DDD	1+AXX+4N(numbers)*	(1)+ABX+4N†‡§
Home area special toll	0+AXX+4N	0+ABX+4N
Foreign area DDD	1+*AXX*+AXX+4N‖	(1)+*A0/1X*+ABX+4N
Foreign area special toll	0+*AXX*+AXX+4N	0+*A0/1X*ABX+4N

* A represents 2 to 9; X represents 0 to 9.

† (1) Indicates that the prefix may be used in common control areas but must be used in SXS areas.

‡ In the initial phase, existing type access codes, such as 112, may be retained. Ultimately access codes other than 1 and 0 should be used only where the switching system design requires their use.

§ All-number calling is now recommended; therefore, ABX represents the two letters of the office name followed by a number or three numbers of the recommended all-number office designation. For the present the first two office digits are limited to 2 to 9.

‖ Italicized codes are area codes

Paper 60-1210, recommended by the AIEE Communication Switching Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Chicago, Ill., October 9–14, 1960. Manuscript submitted June 13, 1960; made available for printing August 5, 1960.

O. Myers is with the Bell Telephone Laboratories, Inc., New York, N. Y.

Table II. Capacity of New Numbering Plan

Initial		
Area codes:	A1/0X	152
Office codes:	ABX (Names)	540
	AXX* (All-Number Calling)	640
Ultimate		
Area codes:	AXX	800
Office codes:	AXX	792†

* X in this case represents the digits 2 to 9. When the new plan is implemented it will represent 0 to 9 and 792 office codes will be available.

† Eight codes in the A-1-1 series are reserved for service codes in common control cities.

Service codes: No change
Operator code: "0"
AHAD permitted: Requires an arrangement to determine whether the number includes an area code.

growth will cause an increase in telephone numbers and, therefore, in office and area codes.

New services will increase the demand for numbers and codes considerably. Today, we need 117 area codes. Most of the increase from the original 86 is due to unanticipated growth, but several area codes were also added for dialing to points beyond the continental United States and Canada such as Hawaii, Alaska, and Puerto Rico. Other new services are also having an effect which will increase as times goes on. Direct inward dialing to PBX (private branch exchanges) is one such service. This requires PBX extensions to have standard 7-digit numbers and results in a telephone number per extension rather than in one per PBX trunk.

Another new service that will require telephone numbers is data switching over the telephone network. Each data station will require one or two numbers, so additional numbers, office codes, and area codes will be added because of this service. Still other uses for area and office codes will be generated by paging systems, mobile systems, military systems, and extension of dialing to additional points.

Requirements of the New Plan

The new plan was devised to meet the following requirements:

CAPACITY

Additional capacity is a basic requirement. Capacity could be extended for a few years beyond the expected exhaust date with relatively modest changes in the present plan but experience with temporizing has been bad. Underestimations have led to difficulties in the past. It seems sensible to set up a large capacity plan that will last well into the next century. The eight-hundred area codes provided by the new plan are considered adequate. This number of codes is obtained by using eight digits, 2–9, in the first code position and ten digits, 1–0, in the second and third positions.

While we have been thinking primarily in terms of capacity that would be adequate for North America, we have not been unaware of the ultimate desirability of a world-wide plan. The door has been left open for arrangements such as doubled prefix digits which will allow the proposed plan to grow into a world-wide plan.

A PROGRAMMABLE PLAN

A plan that would require a simultaneous nation-wide switch from the existing to the new would be completely impractical. Apart from the large costs of such a cutover, the co-ordination of work in thousands of central offices could not conceivably be handled without major difficulties and service interruptions. The plan adopted can be implemented in easy stages.

NO MODIFICATIONS OF CUSTOMERS' EQUIPMENT

The chief objection to modifying equipment, such as telephones on customers' premises, is cost. From the standpoint of features certain modifications of telephones are attractive One which has been discussed is the use of specially designated buttons on dials and on keysets proposed to replace dials, which customers would push instead of dialing prefix digits such as "0" or "1." Other changes have also been suggested, but costs have prevented adoption of any of these ideas as a requirement of the new plan.

COSTS

Reasonable over-all costs are obviously a necessity. It is also desirable that money be expended only where needed and at the time a change in plan is needed.

ABBREVIATED HOME AREA DIALING (AHAD)

The new plan recommends the prefix "1" or "0" followed by the 7-digit number for calls within the home numbering plan area. At the time when area code shortages forces the use of the same code for both an area and an office, arrangements will be needed to determine whether the first three digits represent an area code or an office code.

CUSTOMER ACCEPTANCE

It was felt that a plan that was essentially an extension of the present plan would be more readily accepted by users than a radically new plan. Such a plan would also be most desirable from a customer education standpoint.

Description of the Plan

The numbering plan adopted will be implemented in three stages:

1. A period during which area codes of the present type will be employed.
2. A period during which ABO codes may be used to augment present-type area codes.
3. The ultimate stage when 800 area codes of the AXX-type will be available. At this time 792 AXX-type office codes will also be available with ANC (8×10×10 minus the eight A-1-1 type service codes).

DIALING PROCEDURE PRIOR TO AREA CODE SHORTAGE

Common control areas. Regular (sent-paid station-to-station) foreign area DDD calls will be dialed as at present. Home area DDD calls will be either 7 or 10 digits, depending on the preference of the operating telephone company.

Customer-dialed special toll calls such as person-to-person and collect will require the prefix "0" (zero) which will be followed by seven or ten digits as in the case of regular DDD calls.

For assistance calls, "0" alone will be dialed. Since special toll calls may use the prefix "0", an appropriate circuit will wait where the prefix "0" is used to see if further digits are forthcoming. Service codes will be of the present form, namely A-1-1.

At the option of the operating telephone company, the prefix "1" may be adopted at an early date for regular DDD calls to act as a barrier against reaching toll points in error on local-intended calls.

Step-by-step or mixed areas with rotary dials. For regular DDD calls, the toll office access code "1" followed by seven or ten digits is recommended to simplify dialing.

Customer-dialed special toll calls should use the prefix "0" followed by seven or ten digits.

"0" will still be used. For assistance operator calls when customer dialing of special toll calls is inaugurated with the "0" prefix, existing service codes of the form "11X" will be retained.

DIALING PROCEDURES IN EARLY STAGES OF AREA CODE CONGESTION

When the area code capacity of the present numbering plan is exhausted, unused office codes of the form ABO may be used for area codes. These can provide additional area codes at the expense of office code capacity which would be correspondingly reduced.

ULTIMATE SITUATION

In the mid-1970's additional area code capacity may be needed. The "1" access code for regular DDD calls will then be standard. Area codes will be of the form AXX of which there are 800. There will be a maximum of 792 AXX-type central office codes. It will be desirable and probably essential to adopt all-number calling before office codes of the A1X and AOX type are used.

For special toll calls dialed by customers but still requiring operator assistance such as person-to-person, the prefix "0" will continue to be used.

Assistance operator calls will continue to use "0" alone. No change is contemplated in service codes.

The same procedures will be recommended for common control, step-by-step, and mixed areas. However, no one will be required to adopt all the elements of the proposed plan, particularly with equipment where parts of the plan might not be feasible. Thus, 2- and 3-digit access codes may be used with systems which require them although they are not recommended. A two or three digit access code followed by the prefix "0" for special toll calls may also be used.

Abbreviated home area dialing will be used. Thus home-area toll calls will have the form 1 AXX XXXX or 0 AXX XXXX. It will be necessary to provide a feature to distinguish between eight and 11 digit calls where conflicts exist between area and office codes.

Special barrier codes are not a part of the plan. The access codes "1" and "0" are in themselves barriers against reaching toll points in error, and once they have been adopted their use will be mandatory.

Equipment Modifications Required

All existing telephone switching systems, both local and toll, with the possible exception of step-by-step intertoll equipment, will need to be modified. It is too early to describe the modifications required in detail, but some general observations can be made on the nature of the changes.

All common control local systems must be modified to enable them to do the following:

1. To register the prefix "1" or "0" and to signal the translating equipment which prefix has been received. The prefix will not be sent to the toll office.
2. To check after an initial "0" to see whether more digits are forthcoming. If only "0" is received the call will be routed to the operator.
3. To route calls selectively on the basis of the prefix and the code which follows.
4. To determine whether an 8- or 11-digit call is intended in areas using AHAD.
5. To provide translation capacity for up to 800 area codes and 792 office codes.
6. To determine whether a code is used both as an area code and an office code where AHAD is provided.
7. To block toll calls which are received without a prefix.

The step-by-step local system will need to have special trunks from the "1" and "0" levels. The "1" level trunk will cause an immediate seizure of a trunk to the toll office and a fast seizure of a sender or register at that office. If the second digit is 2 to 9, it and succeeding digits will be repeated to the toll office equipment and stored there. However, if the second digit is a "1," it will be absorbed, the toll office trunk will be immediately released, and the auxiliary selector will accept the third digit and route to a service trunk.

The "0" level trunk will be arranged to distinguish between prefix "0" toll calls such as person-to-person and collect, and "zero" operator calls.

At the toll office some of the important new features are as follows:

1. Additional translation capacity for up to 800 area codes and 792 office codes.
2. Arrangements for determining whether a seven or ten digit number is intended.
3. New trunks for the "0" prefix calls.

Some of the work mentioned has already been started. The simple access code "1" arrangement has been developed for step-by-step and no. *5* crossbar, and will soon be in use in many locations.

The prefix "0" arrangements are also in use in a few locations now arranged for customer dialing of person-to-person toll calls. Many more will use it as they install equipment for customer dialing of calls requiring operator assistance.

A Transistorized Cable Carrier Repeater

C. G. GRIFFITH
ASSOCIATE MEMBER AIEE

THIS PAPER describes a repeater which was designed for a medium-haul cable-carrier system. It is called the Lenkurt type *4587A* repeater unit. By use of transistors, power requirements are minimized. By the use of conservative design techniques, a high degree of reliability is expected.

Since this subject may be of interest primarily to those in telephone operating companies and to circuit engineers, emphasis will be directed at (1) applications of the repeater as related to the basic problem, and (2) some design features and performance of the circuitry.

A repeater for the type *45BN* cable carrier system was described in an earlier paper.[1] Because of the continuing usage of this type of repeater, it has been redesigned as a completely transistorized unit to meet demands for reduced power requirements.

A distinct cost advantage in system power-feed is derived from the use of the transistorized repeaters. Since this repeater consumes only 4 watts, a power-feed point may supply a larger number of repeaters than was possible with higher voltage units.

The Problem

The essential problem in cable transmission is that of amplifying signal levels at certain intervals along the cable route. It is necessary to minimize degradation of system performance caused by noise, crosstalk, and sidetone, as a result of adding amplifiers. Since signal amplitudes vary because of temperature effects on the cable and other conditions, it would be desirable to hold repeater output levels constant. Also, high frequencies are attenuated more than low frequencies in cable transmission, causing a large slope

---

Paper 60-1251, recommended by the AIEE Wire Communication Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Chicago, Ill., October 9–14, 1960. Manuscript submitted June 13, 1960; made available for printing September 1, 1960.

C. G. GRIFFITH is with the Lenkurt Electric Company, San Carlos, Calif.

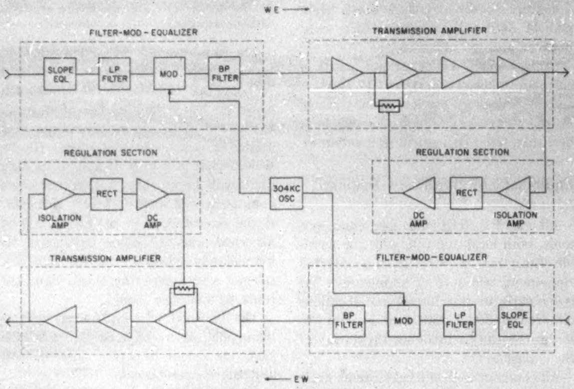


Fig. 1. Block diagram of repeater

to appear across the spectrum. Slope cancellation or appropriate compensation is required to provide a good frequency response for the 24-channel group.

A repeater unit, designed to fill these needs in a cable carrier system, will be discussed.

Description of Operation

The repeater is designed to be used with a 24-channel carrier system. The specific system for which it is intended is the type *45BN* cable carrier system, although the repeater is functionally interchangeable with the type *N* repeater for use in Western Electric "N" and "ON" systems.

The *45BN* is a 24-channel system using single-sideband suppressed-carrier transmission. The 24 channels are transmitted in a range of 40–140 kc (low group) or of 164–264 kc (high group). A terminal that transmits low group will receive high group, or vice versa. A single high-level pilot, 96 kc for low group or 208 kc for high group, is transmitted for regulation. Placed at intervals up to 8 miles, the repeater accepts the incoming frequencies of 40–140 kc or 164–264 kc for both directions of transmission. The main functions provided are frequency translation amplification, and regulation. The repeater dissipates 4 watts by taking a current of 100 milliamperes with a drop of 40 volts. Fig. 1 is a simplified block diagram of the operation.

Frequency translation, or "frogging," has a cancelling effect on the slope of the frequency spectrum caused by increased attenuation with frequency in the cable. Each repeater modulates the incoming group 40–140 kc or 164–264 kc with a 304-kc signal, thus changing a low group to a high group or vice versa. Each channel is thereby inverted as to its order in the group. An incoming negative slope of 8 db (decibels) is transposed to an output positive 8-db slope. The groups are interchanged at successive repeater points as shown in Fig. 2. Each repeater thereby equalizes one-half of the cable section on either side of the repeater, leaving the carrier terminals to provide 8 db of equalization for the half-cable section on each end to complete the system equalization. Small additional cable variations can be compensated for by a variable slope equalizer in the repeater. Each repeater includes a variable-gain amplifier which provides an over-all repeater gain of 46 db at the normal control setting. This gain figure was selected by considering the approximate maximum loss in an 8-mile section of 19-gage cable. To compensate for varying input levels created by temperature effects on the cable, and for other influences, regulation is provided. Output levels will remain constant within 1 db over an input range of ±8 db.

Other Features

Each repeater includes necessary equipment for two directions of transmission. Both circuits may be used to go from low group (40–140 kc) transmission to high group (164–264 kc) or from high to low. Two coaxial connectors are used in each direction of transmission to connect the repeater for the right mode of operation.

Time delay is normally used in the regulator to allow system alarms to function in the event of complete failure of transmission. A strap option makes it possible to have operation without time delay if it is desired.

A gain control is provided to enable adjustment of the output-level operating point over a range of 12 db.

A slope switch with 0, 2, or 4 db of additional slope equalization is provided.

Repeater Circuitry

Filter-Modulator-Equalizer

The incoming signals enter a subassembly called the filter-modulator-equalizer. It is composed of slope-correction networks, a low-pass filter for the 40–140-kc group, a modulator, and a 164–264-kc band pass filter, in that order.

Since the low group will have approximately 4 db more slope than the high group, a slope equalizer is needed to make the frequency-frogging operation successful. To accomplish this, 2 db of slope is removed from the received low groups and 2 db of slope is added to the received high groups.

To allow for minor deviations in slope resulting from unequal lengths of repeater sections, two additional networks are included in the slope-equalizer section of the filter-modulator-equalizer. A 3-position switch on the front of the panel will set the desired condition of either 0, 2, or 4 db of additional slope. Equal loss is

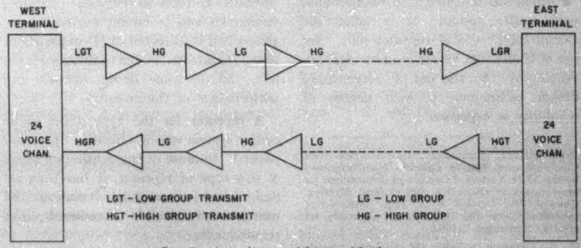


Fig. 2. Interchange of low and high groups

Vol.
Pt. 1

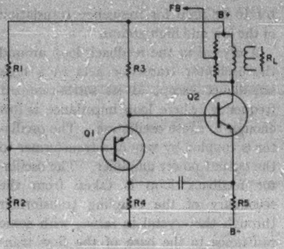


Fig. 3. Schematic of compound transistor

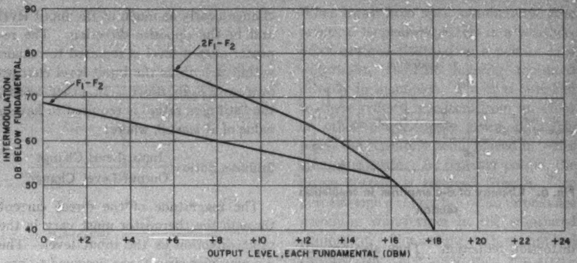


Fig. 4. Intermodulation distortion

given at the pilot frequency for the three positions so that the output level of the repeater is not disturbed by making an adjustment.

The low-pass and bandpass filter design requirements are based on rejection of unwanted out-of-band tones, repeater sing suppression, crosstalk suppression, amplifier unloading, and frequency response. In each filter, a rejection notch is placed at the carrier frequency of 304 kc in order to reduce carrier leak.

The Appert Modulator[2] is used because of its low-distortion characteristics.

A low-group spectrum first passes through the slope equalizer and then through a low-pass filter for the 40–140-kc band. It then modulates a 304-kc signal which is generated within the repeater. A 164–264-kc bandpass filter selects the lower sideband of the products, and from there the high group enters the preamplifier.

An important feature of the filter-modulator-equalizer is that it is bilateral. By interchanging the input-output terminal connections it can be used to receive low group and pass the resulting high group on to the amplifiers, or to receive the high and send on low group.

The insertion loss of the filter-modulator equalizer is 13 db.

TRANSMISSION AMPLIFIER

The transmission amplifier consists of four stages: a preamplifier, a regulation stage, a driver stage, and a power-output stage.

The preamplifier is a common emitter configuration using a drift transistor. After the preamplifier, the remainder of the circuit could be called a closed-loop servomechanism, where a portion of the repeater output is rectified and amplified in the regulator circuitry and used to control the gain of the regulation stage in the transmission amplifier.

The second or regulation stage employs a drift transistor in a common emitter circuit with its gain controlled by a shunt-feedback resistance around the stage. For fixed-gain operation a resistor is used in the feedback circuit, but for the normal regulation mode, a directly heated thermistor is used. The thermally sensitive resistor is controlled by the d-c output of the regulator section.

The third stage, also a common emitter, is a driver for the output power amplifier.

The power-amplifier stage is a compound transistor circuit using a drift transistor and a 4-watt n-p-n silicon power transistor. By the use of a compound circuit, it is possible to obtain low nonlinear distortion and reduce the sensitivity of the amplifier performance to variations in the power transistor.

The "compound transistor"[3] shown in Fig. 3 is a special case of base current feedback.[4] Its configuration is similar to the Darlington compound except that the collector of *Q*1 is returned to the emitter of *Q*2 rather than to the collector of *Q*2. A grounded emitter version of the compound transistor was chosen as it has greater power gain and requires less current from the driver stage because of its high input impedance. Intermodulation distortion of the repeater is shown in Fig. 4.

In order to describe the manner in which distortion reduction has been accomplished, the different forms of distortion will be discussed. These forms of distortion in a transistor, or in any active device, are derived from nonlinearities in: (1) voltage gain, (2) current gain, and (3) input impedance.

First, the notation will be introduced and the pertinent equations given. Consider an active device that is inserted between a resistive load and a source with an internal resistance as shown in Fig. 5. Starting with the relationships $E_1 = I_1 Z$ and $E_2 = I_2 R_L$, the second expression can be divided by the first, giving

$$\frac{E_2}{E_1} = \frac{I_2 R_L}{I_1 Z}$$

Let voltage gain $= V = E_2/E_1$ and current gain $= I = I_2/I_1$; then

$$V = I\frac{R_L}{Z}$$

or

$$R_L = \frac{VZ}{I}$$

Since R_L is a fixed independent linear quantity, the total nonlinearities in VZ/I must cancel. It is possible for one of the quantities on the right-hand side to be linear with the other two nonlinear, but it is not possible for only one to be nonlinear. Therefore, if the current gain I is linearized, the output distortion will decrease as the input impedance Z is made linear. Also, if the voltage gain V is linear, the output voltage will be undistorted when the amplifier is driven from a zero-impedance voltage source. This suggests that the output distortion will decrease as the source impedance Z_0 is decreased.

In the repeater power amplifier, compounding has the effect of making current gain, I, more linear since the input impedance is increased by the addition of resistance between the emitter of *Q*2 and ground. The bias resistors will tend to make the input impedance more linear. The addition of a fixed linear high admittance to a nonlinear low admittance makes this possible. Distortion is further decreased by using a source impedance, Z_0, which is low compared with the input impedance of the compound stage.

*R*1 and *R*2 make up a parallel admittance combination which contributes to

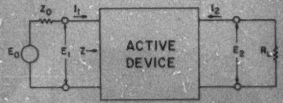


Fig. 5. Terminology associated with 2-port device

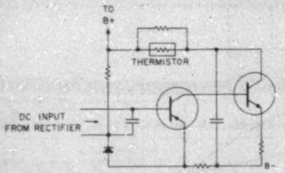


Fig. 6. Circuitry of d-c amplifier in regulation section

distortion reduction. They also provide a stiff base voltage for *Q*1. *R*3 is used to stabilize the direct current of the first transistor's emitter since the emitter current depends on both the constant current through *R*3 and also the base current of the second transistor which varies among units.

The collector of the first transistor is connected to the emitter of the second by a large capacitor which offers a low impedance to signal currents. A type of bridge feedback is used to set the output impedance, further reduce distortion, and stabilize gain.

The over-all gain of the transmission amplifier is 59 db.

REGULATION SECTION

To accomplish regulation, a portion of the output voltage is taken from a tap on the primary winding of the output transformer. This voltage is rectified and amplified and the direct current used to control resistance in the feedback thermistor of the transmission amplifier's second stage.

An isolation amplifier is used to decouple the rectifier distortion from the repeater output. The rectifier output voltage is compared with a reference diode voltage. The difference voltage is the input to a 2-stage d-c amplifier shown in Fig. 6.

In determining the regulator characteristics, the first requirement is that the gain of the regulating amplifier must change nearly as much as the input level and in the opposite direction. The repeater output level is required to remain within ±1 db as the input level changes over a ±8-db range. In other words, the "stiffness ratio" is required to have a value of at least 8, where:

$$\text{Stiffness Ratio} = \frac{\text{Input-Level Change}}{\text{Output-Level Change}}$$

The magnitude of the direct current through the thermistor must vary in the same direction as the input level. The d-c amplifier circuit was designed to produce the specific range of current needed. Since the output range is required to be less than 2 db, this 2-db change in bias at the d-c amplifier is required to produce a range of output current that will cause a gain variation of 16 db in the regulating amplifier.

Regulation time-delay is accomplished by connecting a tantalum capacitor between the emitter of the d-c amplifier's second stage and the base of the first. This acts as a feedback path. An instantaneous change in the output level of the repeater causes an instantaneous change in input voltage to the d-c amplifier. The capacitor prevents any instantaneous change in the output of the d-c amplifier. Thus a sharp drop in input level is not immediately compensated for and the repeater output level also drops for the same amount. Recovery of the output level by regulation action is slow enough to permit operation of system alarms. This provision is necessary because regulation action tends to maintain a high-output level even with a complete loss of input signal. System alarms are set to actuate for a certain decrease in signal power received at the terminal.

304-KC OSCILLATOR

The 304-kc oscillator in Fig. 7 is a crystal-controlled transistor-feedback oscillator. The purpose of the oscillator is to provide a high-level tone for the modulator to be used for frequency translation of the low and high groups.

A crystal in the feedback loop around the oscillator transistor acts as a high impedance except at its series-resonant frequency, where loop impedance is low enough to cause oscillation. The oscillator is coupled by way of a transformer to the output power amplifier. The oscillator feedback loop is taken from the secondary of the coupling transformer through the crystal in series, with some resistance to the base of the first transistor. The resistance includes a negative-temperature coefficient-resistor to compensate for variations in crystal resistance with temperature.

A tank circuit at the transformer primary prevents crystal operation in a spurious mode, decreases distortion, and provides fine tuning. A level control is included in the power amplifier.

Performance Characteristics

The *4587A* repeater unit has a maximum insertion-gain deviation of 0.3 db across the 40–140-kc band and from 164–264 kc. The repeater fixed-gain is 46 db. In the normal-regulation mode of operation, output level may be varied ±6 db by a screwdriver control on the faceplate.

For input variations of ±8 db, the output level remains constant within less than ±1 db. The input impedance is 600 ohms and the output, 130 ohms. (In the associated shelf equipment, the 130-ohm cable impedance is transformed to 600 ohms for the repeater input.) Upon loss of signal, time delay prevents immediate transition of the repeater amplifier to maximum gain. After a sharp decrease in input signal, there is a lapse of at least 7 seconds for a 1-db increase in insertion gain.

Very low sidetone has been obtained. With the two directions of transmission, called West-East (W-E) and East-West (E-W), a tone injected in the West-East

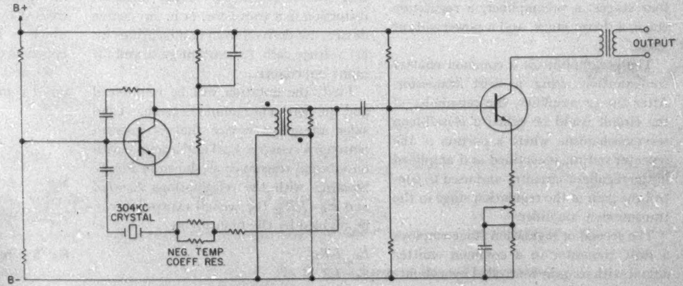


Fig. 7. Schematic of 304-kc oscillator

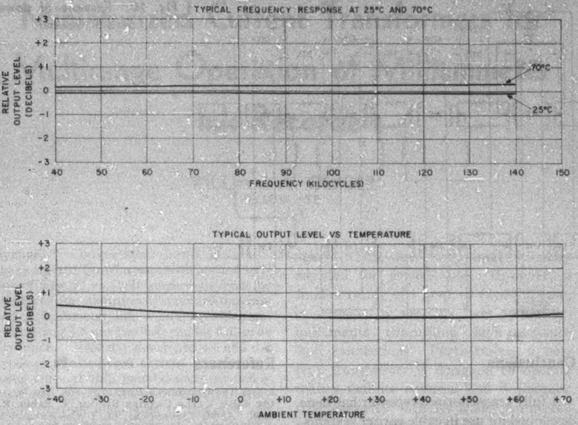


Fig. 8. Temperature effects on output levels and frequency response

direction of transmission is more than 60 db down at the East-West output, and likewise for East-West to West-East.

The repeater has shown insensitivity to extreme temperatures in numerous tests, which included temperature cycling from −40 to +70 C (degrees centigrade) and also prolonged operations at abnormally high temperatures. Performance is not degraded under these conditions, as indicated in Fig. 8.

Physical Characteristics

The repeater unit is the heart of a complete repeater installation. It is designed for plug-in mounting in the Lenkurt line equipment unit. The line equipment unit, in turn, may be mounted either in a repeater shelf or in a gastight container to be placed in an underground location. Also, the repeater only may be mounted in a repeater mounting-adapter for a Western Electric mounting bracket.

The repeater unit occupies a space 8 inches high, 10 inches long, and $3^{5}/_{8}$ inches wide. Its weight is $6^{3}/_{4}$ pounds. The circuitry for both directions of transmission is included. The two filter-modulator-equalizers are wired in units mounted on top of the unit and the two amplifier-regulator circuits are mounted on phenolic cards, one on each side of the unit. One card is a mirror image of the other. The oscillator circuitry is mounted on a card across the back of the unit, while on the front are coaxial connectors for selecting low-high or high-low operation. Below the connectors are two selector switches for additional slope correction, screwdriver controls for setting the output-level operating point, and a metering jack which provides appearance of W-E output level, E-W output level, repeater-unit supply voltage, and the local oscillator level. A view of the repeater unit mounted on a line equipment unit is shown in Fig. 9.

Power-Feed Considerations

Sometimes repeater locations conveniently coincide with a telephone office or a repeater hut needed for other systems. When away from such installations, it is practical to place repeaters in a weatherproof cabinet mounted on a pole or in a watertight, gastight housing mounted directly in manholes. Power is supplied to these remote repeaters from an office or other power-supply point by cable pairs, the repeaters being placed in series along the route. At repeater points, the direct current is received from a cable pair through a center tap on an impedance matching transformer in the associated equipment. The return path is supplied through the cable pair used for transmission in the opposite direction.

A 43-volt zener diode is connected in parallel with each repeater across the 40-volt supply; thus, when a repeater is removed from the line, the voltage across the other repeaters will remain constant.

When repeaters are spaced at 8-mile intervals using 19-gage cable, a total of seven repeaters may be powered from a single supply point using two 130-volt d-c supplies connected together in series. This method is illustrated in Fig. 10. In determining voltage drop in the cable sections, no. *19* American wire gage cable simplex resistance at 43 C is used with a repeater spacing of 8 miles.

System Length

Repeater design considerations, which include noise, intermodulation, and sidetone, are based on a system length of 30 repeaters.

Although the repeater was designed to amplify signal levels that have been at-

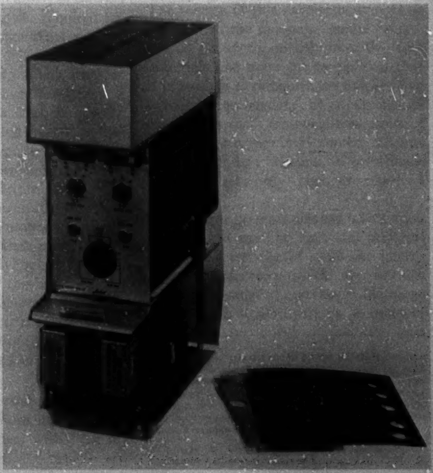


Fig. 9. View of repeater unit and associated line equipment

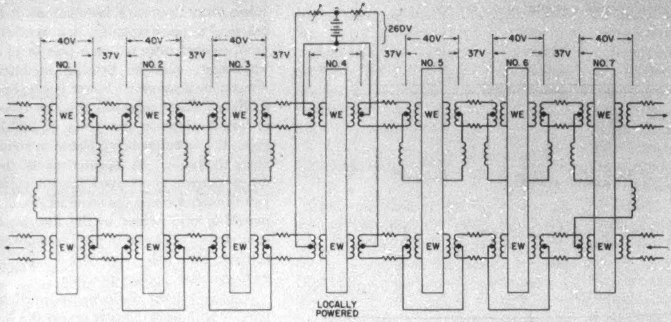


Fig. 10. Example of power feed to repeater

tenuated by 8-mile sections of 19-gage cable, shorter cable lengths may be necessary for various geographical reasons. Also 13-gage or 16-gage cable may be used. Under these circumstances, span pads and artificial lines are available. Receive pads of 0 to 30 db in 2-db steps and transmit pads of 4, 8, 12, 16, or 20 db may be used. Artificial lines of 1, 2, and 4 miles are available for the receive or transmit side.

Conclusions

A fully transistorized repeater has been developed for use in cable carrier systems. Some of the most important advantages include low sidetone, low distortion, reliable operation obtained by use of conservatively rated solid-state devices and low-power consumption. The last is of prime importance in view of high costs for maintaining power-supply facilities.

References

1. A Transistorized Repeater for Use With the 45BN Cable Carrier System, **V. Babin, R. Fish.** *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 77, Mar. 1958, pp. 41–49.

2. Frequency Changing Circuit, **K. E. Appert.** *U. S. Patent no. 2,545,250*, Mar. 13, 1951.

3. A new Compound Transistor, **J. A. Stewart.** *Proceedings*, Institute of Radio Engineers, New York, N. Y., Apr. 1958, pp. 18–20.

4. Base Current Feedback in Power Amplifier Design, **F. S. Boxall.** *Wescon Convention Record*, Institute of Radio Engineers, pt. 2, 1957, pp. 20–26.

Miniaturized Current Transformers for Multirange Operation of Milliammeter Ink Recorders

WILLIAM A. GEYGER
FELLOW AIEE

Synopsis: Conventional forms of multirange current transformers with a rated output of 5 or 10 va (volt-amperes) cannot be employed for operation of standard moving-coil milliammeter ink recorders which require about 1.5 mw (milliwatts) for full-scale deflection. Recently developed a-c and d-c types of miniaturized current transformers using cores of thin molybdenum-permalloy tape and special construction (extremely small cross-sectional area of the core and very large number of turns of the secondary winding) make multirange operation of such recorders possible.

This paper presents detailed design information and performance characteristics of small-size instrument transformers, which are suitable for measurement of currents in the range of 0.1 ma (milliampere) to 10 amp (amperes) with an accuracy of ±0.3 to 0.5%. High a-c and d-c voltages up to about 10,000 volts may be measured by employing the current range 0–1 ma and adding a series resistor (1,000 ohms per volt) corresponding to the multiplier resistor of an ordinary voltmeter.

The paper also describes simple differential-type test circuits for high-precision measurement of actual current ratio and percentage current-ratio error. It presents oscillograms which characterize the dynamic performance of small-size d-c transformers under half-cycle-response operating conditions. With these conditions, the load current is not only linearly proportional to the direct control current, but to some extent will even have similar oscillographic variation when this control current varies, owing to a superimposed a-c component.

THE application of nickel-iron-alloy core materials to instrument transformers makes it possible to reduce the imperfections of such transformers to the minimum and to approach very closely the ideal operating conditions. By providing means for varying the turns ratio of ring-core current transformers, e.g., a tapped primary or a cable primary threaded through the window opening, measurements in the laboratory and test room can be made over a wide range of currents with a high degree of accuracy and with the use of a minimum amount of expensive equipment. As a result of standardization on instrument transformers, the normal rated secondary current for current transformers is 5 amp, a figure which has been internationally adopted for general use. However, a smaller rating of 1 amp, or even 0.5 amp, is occasionally used to meet special requirements concerning high-resistance lead connections. Furthermore, a current transformer is designed to operate with a definite external-load impedance (the ammeter with its leads) connected to the secondary terminals. The rated external load impedance is the impedance value with which the current transformer can be loaded, so that with the rated secondary current flowing, the actual error values do not exceed the limits given by the respective specifications.

In view of the properties and advantages of standard d-c milliammeter permanent-magnet moving-coil ink recorders, recent research and development work has been devoted to the application of similar current-transformer techniques for multirange operation of such recorders and panel-type milliammeters. Of course, the conventional forms of precision-type current transformers with a rated output of 5 or 10 va cannot be employed for this purpose, because the power required for full-scale deflection of a standard moving-coil ink recorder (e.g., Esterline-Angus type, 0–1 ma) is 1.5 mw. Thus, the specific problem consisted of developing a miniaturized form of a 60-cycle current transformer with a rated output in the order of milliwatts for multirange operation of such milliampere ink recorders. An additional requirement is, of course, to provide any desired current ratio, within certain limits, with full ampere-turns and identical accuracy on each range. This can be achieved simply by tapping the primary winding. In this way, by the choice of a suitable turns ratio it is possible to arrange for the readings to be taken at the upper part of the scale where the precision is highest, for all values of the primary current.

A similar situation exists with regard to multirange operation of milliammeter moving coil-ink recorders with "d-c instrument transformers"[1–4] which are based upon the inherent current-transformer properties of ordinary saturable-reactor circuits with nickel-iron-alloy core material.[5] These circuits operate in conjunction with a full-wave rectifier and a moving-coil d-c ammeter which indicates the half-cycle average value of the alternating current flowing through the a-c windings of two equally rated saturable-reactor elements. In this case, the problem consisted of developing a 60-cycle-operated prototype design of a miniaturized d-c instrument transformer, similar to its a-c counterpart.[6]

Closely associated with the development of both the a-c and d-c type of these small-size current transformers are special problems concerning high-precision measurement of actual current ratio. In order to measure with a high degree of accuracy the percentage current-ratio error of these transformers, differential-type test methods have been introduced. Further investigations are concerned with the dynamic performance of miniaturized d-c transformers with half-cycle-response operating conditions. With these conditions, the load current is not only linearly proportional to the direct control current, but to some extent will even have similar oscillographic variation when this control current varies, owing to a superimposed a-c component.

This paper reports on the solution of these problems and describes both a-c and d-c types of miniaturized current transformers having a maximum output of about 50 mw and a rated full-load secondary current of 1.0 ma corresponding to full-scale deflection of a standard 1-ma 1,500-ohm moving-coil ink recorder. The success of this development work is based upon the unique magnetic properties of modern nickel-iron-alloy core materials and upon the recognition that it is possible to minimize the magnetizing current requirements by applying an extremely small cross-sectional area of the core and a very large number of turns of the secondary winding, in accordance with the relatively high external-load resistance.

The paper presents detailed information on core and winding data and performance characteristics of small-size instrument transformers which are suitable for measurement of currents in the range of 0.1 ma to 10 amp with an accuracy of

---

Paper 60-806, recommended by the AIEE Recording and Controlling Instrumentation Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting, Atlantic City, N. J., June 19–24, 1960. Manuscript submitted February 15, 1960; made available for printing June 27, 1960.

WILLIAM A. GEYGER is with the U. S. Naval Ordnance Laboratory, Silver Spring, Md.

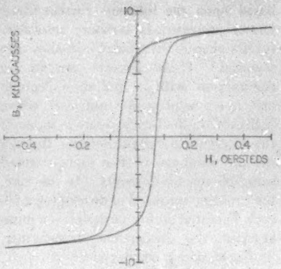


Fig. 1. The 60-cps B-H loop of the Mo-permalloy 2-mil tape cores used in the prototype designs of miniaturized instrument transformers

±0.3 to 0.5%. High a-c and d-c voltages up to about 10,000 volts may be measured by employing the current range 0-1 ma and adding a series resistor (1,000 ohms per volt) corresponding to the multiplier resistor of an ordinary voltmeter. The paper also describes simple differential-type test circuits for high-precision measurement of actual current ratio and presents oscillograms which characterize the dynamic performance of miniaturized d-c instrument transformers under half-cycle-response operating conditions.

Miniaturized Current Transformer for A-C Measurements

DESIGN CONSIDERATIONS

The design of this transformer has been based upon a rated full-load secondary current of 1.0 ma and a rated external resistance of 2,000 ohms. This prototype design, therefore, operates with a resistive external load and a rated secondary output of 2×10^{-3} va. As a result of these quite unusual operating conditions, it is necessary to use a special test circuit to determine the full ampere-turn handling capacity of the current transformer and its performance characteristics showing the percentage current-ratio error as a function of the primary current with the load resistance as a parameter. It is to be noted that the primary current will be measured in terms of the average value of full-wave-rectified secondary current flowing through the moving-coil system of the ink recorder or panel-type milliammeter. Thus, in accordance with the basic principle of rectifier instruments, this arrangement is susceptible to waveform errors which, in general, will depend upon the phase as well as the magnitude of the harmonics that are present. With sinusoidal waveform of the primary current, the ratio of average to rms values is 0.909, and the recorder or panel meter can be calibrated to read rms values of alternating current or voltage.

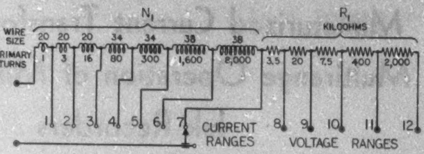


Fig. 2. Auxiliary resistance network with range-selecting switch used in conjunction with the tapped primary winding N_1 of a miniaturized instrument transformer for a-c measurements

With regard to the lowest current range, 0–1,000 μa (microamperes), the magnetizing and iron-loss components of the primary exciting current should be reduced to very small values by the application of nickel-iron-alloy high-permeability core material. In order to reduce the ampere-turn requirements and the current-ratio error to the minimum, the following aspects are considered to be of prime importance:

1. High-permeability core materials having an extremely low coercive force (H_c), e.g., supermalloy type, are definitely preferred in this special application to the grain-oriented nickel-iron alloys in order to minimize the ampere-turn requirements of the current transformer.

2. Although the frequency of the primary current, 60 cps (cycles per second), is low, 2-mil tape cores are considered to be particularly suitable in this application to minimize eddy-current effects.

3. The miniaturized current transformer must be designed to have a very small cross-sectional area of the core and a correspondingly high number of turns of the secondary winding, e.g., 10,000 turns.

4. A sufficiently large core window must be provided to avoid application of undesirable extremely thin wire for the windings, and excessively high values of winding resistance.

PROPERTIES AND DIMENSIONS OF COMPONENTS

In order to minimize the ampere-turn requirements, a Mo-permalloy 2-mil tape core (inside diameter, 1.250 inches; outside diameter, 1.500 inches; tape width, 0.125 inch) is used in the construction of the prototype design. Fig. 1 presents the 60-cps hysteresis loop of this core, which has been automatically recorded by means of a simple dynamic **B-H** loop tracer[7–9] with silicon-diode chopper circuits (Ferrotracer) and a chopper-type *x-y* plotter. The magnetic-core test circuit itself has been designed so that the core sample is tested under sinusoidal flux excitation conditions.

By tapping the primary winding (total number of turns = 10,000 corresponding to the lowest current range, 0–1 ma), the current ratio can be varied conveniently, with full ampere-turns and identical accuracy on each range. When using this lowest current range for measurement of high a-c voltages up to about 10,000 volts, the voltage ranges may be varied by means of a tapped series resistor, or "multiplier," as in the case of an ordinary voltmeter. Actual number of turns of the secondary winding can be changed (10,000, 9,950, or 9,900 turns, respectively, of no. *38*, B & S [Browne and Sharpe] wire) to obtain optimized performance of the current transformer with the rated external-load resistance of 2,000 ohms.

Table I presents the effective numbers of turns of the primary winding, the effective resistances corresponding to the 12 positions of the selecting switch (Fig. 2), and the resultant current ranges, corresponding to the rated-secondary-current value of 1.0 ma, the full-scale deflection of the ink recorder or panel meter. It is to be noted that the ratio

Table I. Effective Turns of Primary Winding N_1 and Effective Resistance Values Corresponding to the 12 Positions of the Selecting Switch

Switch Position No.	Effective Turns of N_1	Current Range, Amperes	Voltage Range, Volts	Effective Resistance, Ohms
1	1	10		0.001
2	4	2.5		0.004
3	20	0.5		0.16
4	100	0.1		4.6
5	400	0.025		19.6
6	2,000	0.005		320
7	10,000	0.001		1,720
8	10,000	0.001	5	5,000
9	10,000	0.001	25	25,000
10	10,000	0.001	100	100,000
11	10,000	0.001	500	500,000
12	10,000	0.001	2,500	2,500,000

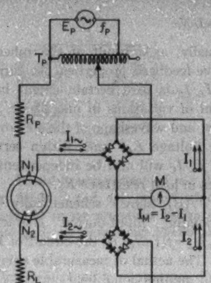


Fig. 3 (left). Differential-type test circuit for high-precision measurement of actual current ratio and percentage current-ratio error of a miniaturized current transformer with the turns ratio $N_1/N_2=1$

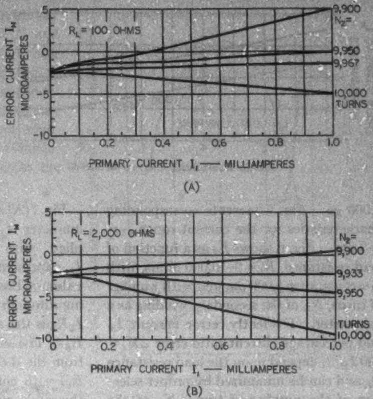


Fig. 4 (right). Error current $I_M=I_2-I_1$ as a function of the primary current, I_1=0–1,000 μa, with number of turns N_2 of secondary winding as a parameter

A—Characteristics with load resistance R_L=100 ohms
B—Characteristics with R_L= 2,000 ohms

of conversion of the current ranges and also that of the voltage ranges, four or five respectively, are selected in such a way that the value to be measured and recorded may be read directly on the linear 50-division scale of the Esterline-Angus recorder. The voltage range may be extended up to about 10,000 volts by employing modern high-resistance components.

High-Precision Measurement of Actual Current Ratio

Since the primary current is measured here in terms of the average value of the full-wave-rectified secondary current, the most important test to be made on this miniaturized current transformer is to find how correctly the secondary current is an exact measure of the corresponding quantity to be measured, i.e., to measure with high precision actual current ratio, preferably with various values of secondary current and external-load resistance. The simple test circuit of Fig. 3 takes advantage of the fact that the turns ratio N_1/N_2=10,000/10,000=1 is available in this special design. Both primary and secondary alternating currents, $I_{1\sim}$ and $I_{2\sim}$, are converted into unidirectional currents, I_1 and I_2, by means of two bridge-type full-wave rectifiers with silicon-diode elements; and these unidirectional currents are superimposed in the moving-coil microammeter M (e.g., −50–0–+50 μa, 600 ohms) so that

$$I_M = I_2 - I_1 = \frac{I_{2\sim}}{k_f} - \frac{I_{1\sim}}{k_f} \qquad (1)$$

where k_f is the form factor, i.e., the ratio between rms value and average value of the alternating currents, $I_{1\sim}$ and $I_{2\sim}$, having identical waveform provided that the current transformer has normal operating conditions with rated values of secondary current and external-load resistance. Thus, the percentage current-ratio error in terms of primary current I_1, i.e., the deviation of the actual current ratio $K_I=I_2/I_1$ from its ideal value N_1/N_2 =10,000/10,000=1, is

$$\Delta K_I = 100\,\frac{I_2 - I_1}{I_1} = 100\,\frac{I_M}{I_1} \qquad (2)$$

This error ΔK_I is positive when, with a given primary current I_1, the actual secondary current I_2 is in excess of its ideal value, i.e., when the ink recorder or panel meter reads too high.

Furthermore, the deviation of actual current ratio I_2/I_1 from its ideal value may be expressed in terms of the maximum value, $I_{1\max}$, of the current range, i.e., in per cent of full-scale deflection of the instrument which measures the average value of secondary current I_2:

$$\Delta K_I\,\frac{I_1}{I_{1\max}} = 100\,\frac{I_2 - I_1}{I_{1\max}} = 100\,\frac{I_M}{I_{1\max}} \qquad (3)$$

The purpose of the test circuit of Fig. 3 is to measure with a high degree of accuracy the percentage current-ratio error, which is due to imperfections of the magnetic core. Obviously, imperfections of the microammeter M measuring the error current $I_M=I_2-I_1$ enter into the measurement as errors of second order. Therefore, it is possible to employ an ordinary panel-type microammeter for measuring ΔK_I with an accuracy of about ±1 to 2%.

Using the test circuit of Fig. 3 with N_1= 10,000 turns and with various numbers of turns N_2, within the range of 9,900 to 10,000 turns, actual magnitude of error current I_M has been measured as a function of the primary current I_1. Fig. 4 shows $I_M=f(I_1)$ with load resistance R_L= 100 ohms (A), and with R_L=2,000 ohms (B) for the lowest current range, I_1=0–

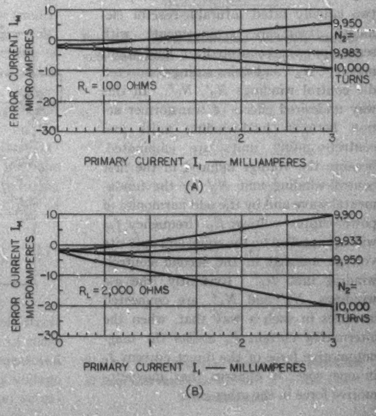


Fig. 5. Error current $I_M=I_2-I_1$ as a function of the primary current, I_1=0–3 ma, with number of turns N_2 of secondary winding as a parameter

A—Characteristics with load resistance R_L=100 ohms
B—Characteristics with R_L= 2,000 ohms

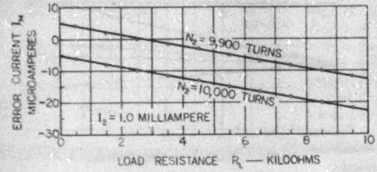


Fig. 6. Error current $I_M = I_2 - I_1$ as a function of load resistance, R_L = 0—10 kilohms, with I_2 = constant = 1.0 ma, and with number of turns N_2 of secondary winding as a parameter

1,000 μa. Fig. 5 presents corresponding characteristics for the current range I_1= 0–3 ma. Fig. 6 shows I_M as a function of load resistance, R_L=0–10,000 ohms, with I_2=constant=1,000 μa, and with number of turns, N_2, of the secondary winding as a parameter. Evidently, error current I_M and percentage current-ratio error $\Delta K_I = 100\, I_M/I_1$ depend upon the load resistance R_L and can be minimized by proper selection of the number of turns, N_2, of the secondary winding (9,900, 9,933, 9,950, 9,967, and 10,000 turns, respectively).

The results of these measurements show that actual current ratio varies within the limits of about ±0.2 to 0.3% with optimum adjustment of the number of turns of the secondary winding. Therefore, this prototype design of a miniaturized current transformer for a-c measurements is suitable for multirange operation of standard moving-coil ink recorders and ordinary panel meters without impairing the accuracy of measurements corresponding to the performance of such instruments (about ±1 to 2% of full-scale deflection).

Miniaturized D-C Instrument Transformer

Basic Saturable-Reactor Circuit

The ordinary "series-connected" saturable-reactor circuit, Fig. 7(A), employs two equally rated saturable-reactor elements (twin-core arrangement) with series-opposing-connected a-c load windings N_L', N_L'', and series-aiding-connected d-c control windings N_C', N_C''. In this way undesired effects of transformer action between load-winding units and control-winding units are eliminated, because the voltage induced in the first control-winding unit N_C' by the fundamental wave and by the odd harmonics of power-supply voltage E_P (frequency f_P) will be cancelled by the equal and opposite voltage induced in the second control-winding unit N_C''. Evidently, the a-c windings, N_L' and N_L'', are connected in series in such a way that, when the alternating current I_N assists the magnetomotive force of the direct current I_C in one core, it opposes that magnetomotive force in the other core.

Fig. 7(A) shows that the moving-coil d-c ammeter A_L (load resistance R_L) is supplied with full-wave-rectified current I_L representing the half-cycle average value of the alternating current I_N flowing through the a-c load windings, and I_LR_L is the average value of load voltage E_L. The control windings are supplied from the d-c signal source (signal voltage E_S) with control current I_C representing the direct current to be measured.

Application of the arrangement of Fig. 7(A) for measurement of direct currents is based upon the current-transformer character (law of equal ampere-turns[10]) which is the result of inherent negative-feedback properties ("intrinsic" negative feedback) of the ordinary saturable-reactor circuit using nickel-iron-alloy core material. That is, the ratio between the average value of the rectified load current I_L and the average value of the control current I_C is substantially equal to the ratio between the number of turns of d-c control windings $N_C' = N_C'' = N_C$, sometimes termed "primary windings," and a-c load windings $N_L' = N_L'' = N_L$, sometimes termed "secondary windings":

$$\frac{I_L}{I_C}=\frac{N_C}{N_L} \tag{4}$$

or

$$I_LN_L=I_CN_C \tag{5}$$

Actually, as a result of the inherent negative-feedback properties, the current ratio I_L/I_C is to a certain extent independent of variations in magnitude, frequency, and waveshape of the a-c power-supply voltage E_P; and, within certain limits, I_L/I_C will also be independent of changes in load resistance R_L.

Fig. 7(B) shows schematically the corresponding input-output characteristic $I_LN_L = f(I_CN_C)$ of the circuit of Fig. 7(A). The actual or measurable average value of unidirectional load current I_L is the resultant of the "exciting-current component" I_{LO}, which is practically independent of the control current I_C, and of the "load component" I_{LC} flowing in response to changes in control current I_C. Equal increments in control ampere-turns, ΔI_CN_C, produce approximately equal increments in load-current ampere-turns, ΔI_LN_L, in the linear region of the characteristic. Thus, since I_{LO} representing the magnetizing current of the saturable-reactor elements can normally be made very small compared with I_{LC}, equation 5 holds to a good degree of accuracy. However, this is true only when the measured load current I_L is expressed in average values over one half-cycle. Therefore, it is essential that a moving-coil instrument with a full-wave rectifier circuit be used. Instruments measuring rms values, such as iron-vane ammeters, will not give the same results.

Specific Problem and Design Considerations

The 60-cycle-operated prototype design of a miniaturized d-c instrument transformer, similar to its a-c counterpart described in the first section of this paper,

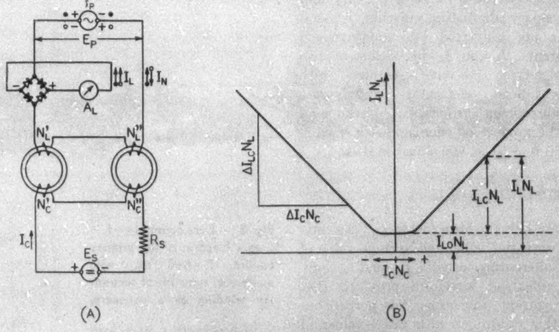


Fig. 7. The series-connected saturable reactor

A—Circuit diagram
B—Input-output characteristic $I_LN_L = f(I_CN_C)$

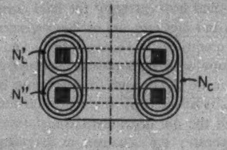


Fig. 8. Toroidal arrangement of a-c load windings $N_L'N_L''$ and d-c control winding N_C employed in the prototype design of a miniaturized d-c instrument transformer

should make it possible to measure direct and low-frequency alternating currents in the range of from 1/1,000 to 10 amp with an accuracy of about ±0.5 to 1%. When using the current range 0–1 ma, a series resistor (1,000 ohms per volt) will be provided for measurement of high d-c voltages up to about 10,000 volts; and a standard 1-ma 1,500-ohm moving-coil ink recorder may be used in conjunction with this arrangement.

Because the maximum load-current value corresponding to full-scale deflection of the recorder is 1,000 μa, the magnetizing-current component of the saturable-reactor elements, Fig. 7(A), should be of the order of 1% of this value, that is, about 10 μa, or less. In order to reduce the exciting-current component I_{LO}, Fig. 7(B), to the minimum, the following aspects, similar to those concerning the miniaturized current transformer for a-c measurements, are considered to be of prime importance:

1. Although the rectangularity of the dynamic hysteresis loop of the magnetic core materials with very low coercive force (H_c), e.g., supermalloy type, is less pronounced than that of the grain-oriented nickel-iron alloys, the low-coercive-force materials are definitely preferred in t is design in order to minimize the magnetizing current requirements of the two saturable-reactor elements.
2. Although the saturable-reactor elements are operated with the low power-supply frequency of 60 cps, 2-mil tape cores are employed to minimize eddy-current effects.
3. The saturable-reactor elements must be designed to have a very small cross-sectional area of the core and a correspondingly high number of turns of a-c windings, $N_L' = N_L''$, Fig. 7(A).
4. A sufficiently large core window must be provided to avoid application of undesirable extremely thin wire for the windings, N_L', N_L'', N_C, and excessively high values of winding resistance.
5. In order to reduce actual magnitude of the magnetizing-current component, I_{LO} in Fig. 7(B), to the minimum, it is desirable to provide an a-c bais circuit[11] fed from the power-supply voltage E_P, Fig. 7(A).

In accordance with common practice in the design of magnetic amplifiers,[12] it is advantageous to apply two individually wound a-c windings, N_L' and N_L'', and only one d-c control winding, N_C, which embraces both magnetic circuits; Fig. 8. With this arrangement of saturable-reactor windings, there is no resultant a-c flux through the common control windings, N_C, and no net voltage of fundamental frequency can appear across the control-circuit terminals.

Properties and Dimensions of Components

Mo-permalloy 2-mil tape cores with magnetic properties characterized by the 60-cps **B-H** loop of Fig. 1 were used in the construction of Fig. 8. The core and winding data may be summarized as follows:

Mo-permalloy 2-mil tape cores: tape width = 0.125 inch; inside diameter = 1.250 inches; outside diameter = 1.500 inches.

A-c load windings, N_L' and N_L'': 10,000 turns each of no. *38* (B&S) wire; resistance, 760 ohms each.

D-c control winding, N_C: 10,000 turns with six taps and sections having different wire sizes, corresponding to the arrangement shown in Fig. 2; total resistance, 1,720 ohms.

The full-wave rectifier bridge circuit consists of four silicon diodes (Texas Instrument Company, no. *606*).

The effective numbers of turns of the d-c control winding, N_C, corresponding to those of the primary winding N_1 of the a-c transformer are indicated in Table I. By tapping N_C in this way, the current ratio of the d-c instrument transformer can be varied conveniently, with full ampere-turns and identical accuracy on each range. When using the lowest current range, 0–1 ma, for measurement of high d-c voltages up to about 10,000 volts, the voltage ranges may be varied by means of a tapped series resistor (multiplier), as in the case of an ordinary voltmeter. Since the ratio of conversion of the current ranges, and also that of the voltage ranges, is four or five, respectively, the value to be measured and recorded may be read directly on the linear 50-division scale of the Esterline-angus recorder. In this case, too, the voltage ranges may be extended to higher values by employing modern types of high-resistance components.

Compensation of Magnetizing-Current Component with A-C Bias Circuit

In the orginal saturable-reactor circuit of Fig. 7(A), the average value of magnetizing current I_{LO} has been measured as a function of the power-supply voltage E_P. Fig. 9 shows $I_{LO}=f(E_P)$ and actual

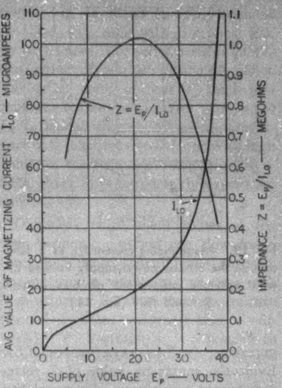


Fig. 9. Magnetizing current I_{LO} and actual impedance $Z = E_P/I_{LO}$ as a function of the 60-cps power-supply voltage E_P

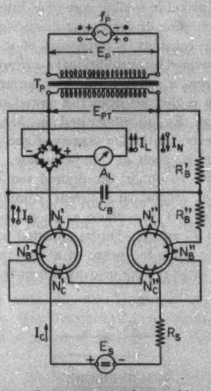


Fig. 10. Series-connected saturable reactor with additional a-c windings N_B', N_B'' acting as bias windings

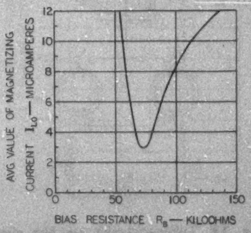


Fig. 11. Magnetizing current I_{LO} as a function of total bias resistance $R_B = R_B' + R_B''$ (C_B = 0.1 microfarad)

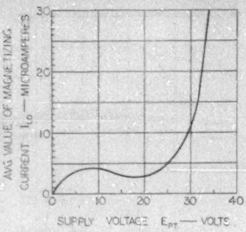


Fig. 12. Magnetizing current I_{LO} as a function of the 60-cps power-supply voltage E_{PT} with optimum adjustment of bias resistors R_B' and R_B'' (Fig. 11)

impedance $Z=E_P/I_{LO}$ as a function of E_P. Evidently, maximum value of Z corresponding to a maximum of permeability occurs with $E_P=21$ volts, and $I_{LO}=20$ μa (stand-by power$=E_P\ I_{LO}=420$ microwatts).

Since the magnetomotive force required by a nickel-iron-alloy core over intervals of increasing or decreasing fluxes is practically constant, a large part of it may be provided by an additional a-c winding[11] fed from a synchronous bias current I_B which is derived from the power-supply voltage E_P, as illustrated in Fig. 10. However, with low-coercive-force core materials, such as Mo-permalloy, it is possible to obtain the desired reduction of the magnetizing-current component I_{LO} only if the phase relationship between I_B and E_P can be adjusted to the proper value by means of a correspondingly rated reactor or capacitor in the a-c bias circuit. In this case, adjustment of the respective components of the a-c bias circuit must be made so that the fundamental waves of I_{LO} and I_B are in phase, in order to bring I_{LO} to the minimum.[11]

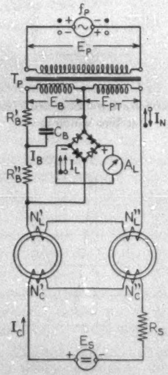


Fig. 13. Differential-type a-c bias circuit which eliminates the additional a-c windings N_B', N_B'' (Fig. 10) of the saturable-reactor elements

Fig. 10 illustrates the application of such an a-c bias circuit in the series-connected saturable-reactor circuit of Fig. 7(A). The series -opposing- connected a-c bias windings N_B', N_B'' have, when using the dimensions of components described in the preceding section, 1,000 turns each of no. *38* (B&S) wire (resistance, 90 ohms each). Bias resistors R_B', R_B'' and shunt capacitor C_B must be properly rated to obtain optimum value of phase displacement between the fundamental waves of bias current I_B and secondary voltage E_T of the power-supply transformer T_P.

Fig. 11 shows the average value of the magnetizing-current component I_{LO} as a function of total bias resistance $R_B=R_B'+R_B''$ with $C_B=0.1$ microfarad. With optimum adjustment ($R_B'=38{,}000$ ohms, and $R_B''=42{,}000$ ohms), I_{LO} has been reduced to about 3 μa, that is to about 1/7 of its original value without a-c bias; Fig. 9. Fig. 12 shows $I_{LO}=f(E_{PT})$ for the case that the a-c bias circuit, Fig. 10, is provided with optimum adjustment of R_B' and R_B''.

Similar effects can be introduced by applying the differential-type circuit of Fig. 13 with a modified power-supply transformer T_P having a center-tapped secondary winding. This advantageous arrangement eliminates the additional a-c windings, N_B' and N_B'', in Fig. 10, and simplifies the construction of the saturable-reactor elements, Fig. 8. With $R_B'=60{,}000$ ohms, $R_B''=1.1$ megohms, and $C_B=0.1$ microfarad, optimum results have been obtained with this arrangement. Here, again, the magnetizing-current component I_{LO} has been reduced to about 1/7 of its original value without a-c bias (Fig. 9).

HIGH-PRECISION MEASUREMENT OF ACTUAL CURRENT RATIO

The purpose of a d-c transformer is to measure the direct current or low-frequency alternating current (I_C) flowing through the control winding (N_C) in terms of the average value of unidirectional load current (I_L). When measuring voltages, the direct voltage or low-frequency alternating voltage is translated into a corresponding control current (I_C) by means of a series resistor, as in the case of an ordinary voltmeter. Therefore, by far the most important test to be made on a d-c instrument transformer is to find how correctly the load current I_L is an exact measure of the corresponding quantity to be measured, i.e., to measure with high precision actual current ratio I_L/I_C and to determine how I_L/I_C is affected by the condition under which the d-c instrument transformer is to operate; e.g., changes in magnitude, frequency, and waveshape of the power-supply voltage.

In accordance with common practice in the magnetic-amplifier art, it is also possible to perform summation of several direct currents by means of several properly rated control windings, which are isolated from each other. This totalizing type of d-c instrument transformer lends itself readily to the metering of direct currents in interconnected systems at different voltages.

The novel test circuit of Fig. 14, a modification of the previously described d-c potentiometer device,[13] permits high-precision measurement on d-c instrument transformers having the turns ratio $N_C/N_L=1$ with regard to actual current ratio I_L/I_C and the influence of variations in magnitude, frequency, and waveshape of the power-supply voltage E_P, etc. This circuit takes advantage of the fact that the turns ratio $N_C/N_L=10{,}000/10{,}000=1$ is available in this special design. Load current I_L and control current I_C are superimposed in the moving-coil microammeter $M(-50\ldots0\ldots+50$ μa, 600 ohms) so that $I_M=I_L-I_C$. Thus, the percentage current-ratio error in terms of I_C, i.e., the deviation of the

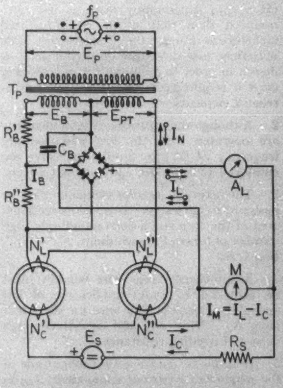
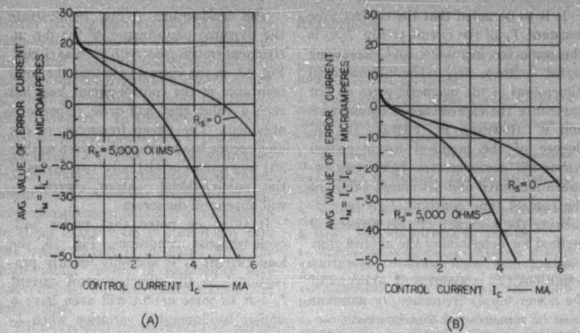


Fig. 14. Differential-type test circuit for high-precision measurement of actual current ratio and percentage current-ratio error of d-c instrument transformers with the turns ratio $N_C/N_L=1$

Vol
Pt.
196



(A)

(B)

Fig. 15. Error current $I_M = I_L - I_C$ as a function of the control current I_C (circuit of Fig. 14)

A—Characteristics without a-c bias
B—Characteristics with a-c bias

actual current ratio $K_I = I_L/I_C$ from its ideal value, $I_L/I_C = 1$, is

$$\Delta K_I = 100\frac{I_L - I_C}{I_C} = 100\frac{I_M}{I_C} \qquad (6)$$

This error ΔK_I is positive when, with a given control current I_C, the actual load current I_L is in excess of its ideal value; i.e., when the milliammeter A_L reads too high.

Furthermore, the deviation of actual current ratio I_L/I_C from its ideal value may be expressed in terms of the maximum value, $I_{C\max}$, of the control-current range; i.e., in per cent of full-scale deflection of the instrument A_L (moving-coil ink recorder) which measures the average value of load current I_L in the output circuit:

$$\Delta K_I \frac{I_C}{I_{C\max}} = 100\frac{I_L - I_C}{I_{C\max}} = \frac{I_M}{I_{C\max}} \qquad (7)$$

The test circuit of Fig. 14 makes it possible to measure with a high degree of accuracy the percentage current-ratio error, which is due to imperfections of the saturable-reactor elements. Because imperfections of the microammeter M measuring the error current $I_M = I_L - I_C$ enter into the measurement as error of second order, an ordinary panel-type microammeter can be used for measuring ΔK_I with an accuracy of about ±1 to 2%.

Employing the test circuit of Fig. 14 with $N_C = 10,000$ turns and the current range 0–6 ma, actual magnitude of error current I_M has been measured as a function of I_C under various operating conditions. Fig. 15 shows $I_M = f(I_C)$ without a-c bias (A), and with a-c bias (B). In both cases, the series resistor of N_C was either $R_S = 0$ or $R_S = 5,000$ ohms. Evidently, error current I_M and percentage current-ratio error $\Delta K_I = 100\ I_M/I_C$ depend upon the actual total resistance of the control-circuit loop: With high values of the series resistance, e.g., $R_S = 5,000$ ohms, I_M is considerably larger than with $R_S = 0$.

Obviously, with small values of R_S, even-harmonic currents can flow freely through the control-circuit loop ("natural"-magnetization conditions), whereas with high values of R_S, these currents are suppressed ("forced"-magnetization conditions).[14–16] Fortunately, it is possible to eliminate this undesirable effect of changes in control-circuit-loop resistance upon I_M by providing an artificial path for the even-harmonic currents, without applying additional d-c windings on the saturable-reactor elements. In the simple arrangement of Fig. 16, a shunt resistor R_N with series capacitor C_N is connected across the control winding N_C. The blocking capacitor C_N prevents direct current from flowing through the shunt resistor R_N. The optimum value of R_N, which is not critical, has been determined empirically by means of the test circuit of Fig. 14.

Fig. 17 shows $I_M = f(I_C)$ without a-c bias (A), and with a-c bias (B). Here, again, the series resistor of control winding N_C was either $R_S = 0$ or $R_S = 5,000$ ohms. The results of these measurements show that the characteristics with $R_S = 5,000$ ohms are quite different from those with $R_S = 0$. However, in both cases, error current I_M is within the limits of about ±10 to 20 μa. Further-

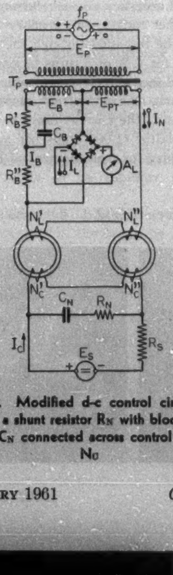
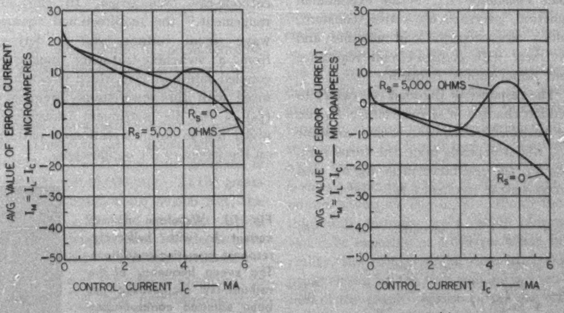


Fig. 16. Modified d-c control circuit employing a shunt resistor R_N with blocking capacitor C_N connected across control winding N_C



(A)

(B)

Fig. 17. Error current $I_M = I_L - I_C$ as a function of the control current I_C (circuit of Fig. 16)

A—Characteristics without a-c bias
B—Characteristics with a-c bias

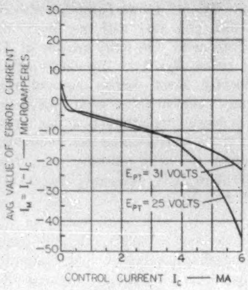


Fig. 18. Error current $I_M = I_L - I_C$ as a function of the control current I_C with 60-cps power-supply voltage E_{PT} as a parameter

more, it is clearly evident from Figs. 15 and 17 that application of an a-c bias circuit, Figs. 10 and 13, is definitely very advantageous.

Similar measurements with the test circuit of Fig. 14 concerning the influence of the usual variations in magnitude and frequency of the power-supply voltage have shown that the error current I_M may change within the limits of about ±30 μa. Fig. 18 shows as an example $I_M = f(I_C)$ with $E_P = 25$ volts and $E_P = 31$ volts, respectively.

DYNAMIC PERFORMANCE

The transient response of the series-connected saturable-reactor circuit, Fig. 7(A), using nickel-iron-alloy core material can be described by simple linear equations.[17,18] When operating in the linear region of the input-output characteristic, Fig. 7(B), the response of the load current I_L to changes of d-c control voltage E_S, power-supply voltage E_P, and load resistance R_L, is an exponential function, governed by a time constant, which varies with the load resistance and inversely with control-circuit resistance and power-supply frequency.

The response of the load current I_L to sudden changes of signal voltage E_S can be described by the control time constant T_C, which depends upon the turns ratio N_C/N_L, the resistance ratio R_O/R_C, and power-supply frequency f_P:

$$T_C = \frac{1}{4f_P}\,\frac{R_O}{R_C}\left(\frac{N_C}{N_L}\right)^2 \text{ seconds} \qquad (8)$$

or

$$T_C = \frac{1}{4}\,\frac{R_O}{R_C}\left(\frac{N_C}{N_L}\right)^2 \text{ cycles} \qquad (9)$$

where R_O is the total resistance of the load circuit and R_C is the total resistance of the control-circuit loop including the internal resistance of the signal source.

It is to be noted that the control time constant T_C of the circuit of Fig. 7(A) is the same for increasing and decreasing signal voltage E_S. The geometrical dimensions of the magnetic cores do not enter into the expressions of equations 8 and 9. However, in practice, the influence of the nonideal magnetization curve will cause certain differences between measured and calculated values of the control time constant T_C. In considering the accuracy of this theoretical method for determining the control time constant for slow-response operating conditions (T_C is about 6 to 30 cycles of the power-supply frequency, or more) it must be remembered that in most practical design problems an approximate value is all that is required.

With regard to fast-response operating conditions, where actual response time is of the order of 1 to 4 half-cycles of the power-supply frequency, experimental methods have assumed increased importance. A simple and convenient method for studying the dynamic performance of d-c instrument transformers and magnetic amplifiers with fast-response operating conditions consists of applying to the control circuit a synchronous unidirectional square-wave signal voltage, which may be derived from a synchronously operated mechanical-switch device[19] or from a magnetic switch[20,21] which consists essentially of a special saturable-reactor circuit producing a synchronous auxiliary current with rectangular waveshape.

The dynamic performance of the miniaturized d-c instrument transformer described in this paper has been studied by means of a modified form of transient analyzer which permits direct visual observation of a stationary pattern on a cathode-ray oscilloscope. In this arrangement,[22] the unidirectional square-wave signal voltage (E_S) is derived from a switching-transistor magnetic-coupled multivibrator which is synchronized with the power-supply frequency (f_P) of the d-c instrument transformer under test to obtain a stationary pattern on the screen of the oscilloscope.

The oscillograms of Fig. 19 illustrate the dynamic performance of the d-c instrument transformer with a-c bias circuit, Fig. 13, in the special case that the total resistance of the control-circuit loop has such a value that half-cycle response is obtained. The sweep frequency of the oscilloscope has been synchronized so that (A) four cycles or (B) only two cycles of the unidirectional output current I_L will appear on the screen.

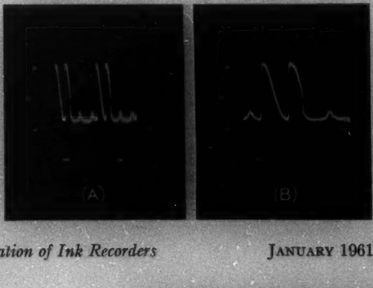
It is to be noted that under half-cycle-response conditions, Fig. 19, the load current I_L is not only linearly proportional to the direct control current I_C but to some extent will even have a similar oscillographic variation when I_C varies, owing to a superimposed a-c component.[23–25]

Conclusions

Recently developed a-c and d-c types of miniaturized instrument transformers for multirange operation of standard permanent-magnet moving-coil ink recorders and d-c panel meters utilize the unique magnetic properties of nickel-iron-alloy high-permeability core materials and the possibility of minimizing the magnetizing-current requirements with special core construction (very small cross-sectional area of the Mo-permalloy 2-mil tape core, and large number of turns of the secondary winding). The design of such transformers having a maximum output power of about 50 mw is based upon a rated full-load secondary current of 1.0 ma. By tapping the primary winding any desired current ratio can be obtained, with full ampere-turns and identical accuracy on each current range. For higher ranges, the primary conductor can be threaded through the window opening for 1, 2, or 5 turns to give the 1-ma rated secondary-current value for 10-, 5-, or 2-amp primary current, respectively. When using the current range 0–1 ma, a series resistor (1,000 ohms per volt) will be provided for measurement of high a-c and d-c voltages up to about 10,000 volts.

The miniaturized d-c instrument trans-

Fig. 19. Waveform of load current I_L with half-cycle-response operating conditions. The sweep frequency of the cathode-ray oscilloscope has been adjusted corresponding to (A) four cycles, or (B) two cycles of the power-supply frequency f_P (60 cps) of the d-c instrument transformer under test (Fig. 13)



Vol.
Pt. 1
196

former having two equally rated saturable-reactor elements with low-coercive-force 2-mil tape cores makes it possible to measure direct currents and low-frequency alternating currents without the use of shunt resistors. In this case, a properly rated a-c bias circuit is used to reduce the magnetizing currrent of the saturable-reactor units to an extremely small value, about 3 μa. A similar arrangement permits measurement of high d-c voltages, up to about 10,000 volts, in such a way that the meter circuit is insulated from the high-voltage circuit, in accordance with conventional a-c instrument-transformer techniques.

High-precision measurement of actual current-ratio error of such small-size instrument transformers using differential-type test circuits make it evident that currents in the range of from 1/1,000 to 10 amp can be measured with an accuracy of ±0.3 to 0.5%. In any case, the primary current flowing through the tapped primary winding or through the single-turn primary will be measured in terms of the average value of full-wave-rectified secondary current flowing through the moving-coil system (range: 0–1.0 ma, 1,500 ohms) of the ink recorder.

References

1. The Transformer for Measuring Large Direct Currents, H. J. Ryan. *AIEE Transactions*, vol. 20, 1901, pp. 169–83.

2. Measurement of Large Direct Currents Over Long Distances (in German), E. Besag. *Elektrotechnische Zeitschrift*, Wuppertal-Elberfeld, Germany, vol. 40, 1919, pp. 436–37.

3. A Simple D-C Instrument Transformer Having Real Current-Transformer Properties (in German), W. Krämer. *Ibid.*, vol. 58, 1937, pp. 1309–13.

4. A New D-C Instrument Transformer for the Measurement of High D-C Voltages, W. Krämer. *Ibid.*, vol. 59, 1938, pp. 1295–98.

5. Single-Ended Saturable Reactor Circuit with Quiescent Current Compensation, R. J. Radus. *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 73, 1954 (Jan. 1955 section), pp. 597–603.

6. A Miniaturized D-C Instrument Transformer, W. A. Geyger. "Special Technical Conference on Non-Linear Magnetics and Magnetic Amplifiers," *AIEE Special Publication T-116*, Sept. 23–25, 1959, pp. 130–53.

7. The Ferrotracer, A Simple Dynamic B-H Loop Tracer With Silicon-Diode Chopper Circuits, W. A. Geyger. *Proceedings*, National Electronics Conference, Chicago, Ill., vol. 12, 1956, pp. 83–97.

8. Magnetic-Switch B-H Loop Tracer, W. A. Geyger. *Electronics*, New York, N. Y., vol. 29, no. 10, 1956, pp. 167–69.

9. Magnetic-Amplifier Circuits (book), W. A. Geyger. McGraw-Hill Book Company, Inc., New York, N. Y., second edition, 1957, pp. 350–69, Fig. 20.10.

10. Magnetic Amplifiers (book), H. F. Storm. John Wiley & Sons, Inc., New York, N. Y., 1955, pp. 77–79.

11. See reference 8, pp. 77–81.

12. *Ibid.*, pp. 27–28, Fig. 2.3(f).

13. Magnetic Amplifiers with Orthonol Tape Cores, W. A. Geyger. *Proceedings*, National Electronics Conference, vol. 6, 1950, pp. 59–67, Fig. 9.

14. D-C Controlled Saturable Reactors and Their Feedback (in German), T. Buchhold. *Archiv für Elektrotechnik*, Berlin, Germany, vol. 36, 1942, pp. 221–38.

15. See reference 10, pp. 64–65, 73–74, 309.

16. See reference 9, pp. 46, 57, 282–83.

17. Series-Connected Saturable Reactor with Control Source of Comparatively Low Impedance, H. F. Storm. *AIEE Transactions*, vol. 69, pt. II, 1950, pp. 756–65.

18. Series-Connected Saturable Reactor with Control Source of Comparatively High Impedance, H. F. Storm. *Ibid.*, pp. 1299–1309.

19. Magnetic-Amplifier Transient Analyzer, W. A. Geyger. *Electronics*, vol. 26, no. 7, 1953, pp. 189–91.

20. Magnetic-Switch Transient Analyzer, W. A. Geyger. *Ibid.*, vol. 29, no. 1, 1956, pp. 150–51.

21. Magnetic-Amplifier-Operated Ink Recorders, W. A. Geyger. *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 77, Sept. 1958, pp. 457–71, Figs. 1, 2, 5, 6.

22. See reference 6, pp. 142–43, Fig. 21.

23. Separately Excited Current Transformer Operating as a Universal-Type Instrument Transformer for Oscillographic Investigations on Alternating Currents with Direct-Current Components (in German), W. Krämer. *Elektrotechnische Zeitschrift*, vol. 60, 1949, pp. 393–95.

24. See reference 10, pp. 167–69, Fig. 10.26.

25. See reference 9, pp. 44–48, Fig. 3.3.

Protection for Underwater Instruments

HAROLD E. EDGERTON
FELLOW AIEE

LLOYD D. HOADLEY
NONMEMBER AIEE

Synopsis: Waterproof cases for instruments, especially cameras, electronic flash lamps, and sonar drivers, are considered for external pressures up to 20,000 psi (pounds per square inch) which is adequate for the greatest known depths in the sea. Equations, charts, and experimentally determined values of failure are given for cylinders, end caps, and windows of various materials. Several types of pressure-resisting electric connections are shown and a pressure-testing facility is described.

THE COMPREHENSIVE article on photography in the ocean by Ewing, Vine, and Worzel[1] has a very complete bibliography covering the work on deep-sea photography from the early work of Boutan in France up to 1946. A series of designs that have proved to be helpful to equipment designers are given in the reference. This paper should be considered an extension of this important contribution.

Those who work in oceanographic instrumentation are required to put their devices in watertight containers for immersion at great depths. The pressure increases 1 atm (atmosphere) for each 32 feet. Therefore, at 2 miles depth (the average of the oceans) the pressure is about 330 atm (5,000 psi). At the greatest known depth, some 36,000 feet, the pressure is 1,130 atm or about 17,000 psi. Case designs must be adequate to withstand these continuous pressures.

The shape of the case is important in the design. A good first impression is made by the sphere shape because of its ability to equalize the forces on the walls, resulting in the lightest wall sections. However, it is difficult to utilize the interior of a sphere effectively: it is more difficult to machine and windows and electric leads disturb symmetry. The next simplest structure is the cylinder. This paper considers the design of cylindrical structures with flat end caps, with commercially available tubes and pipes of several materials in mind.

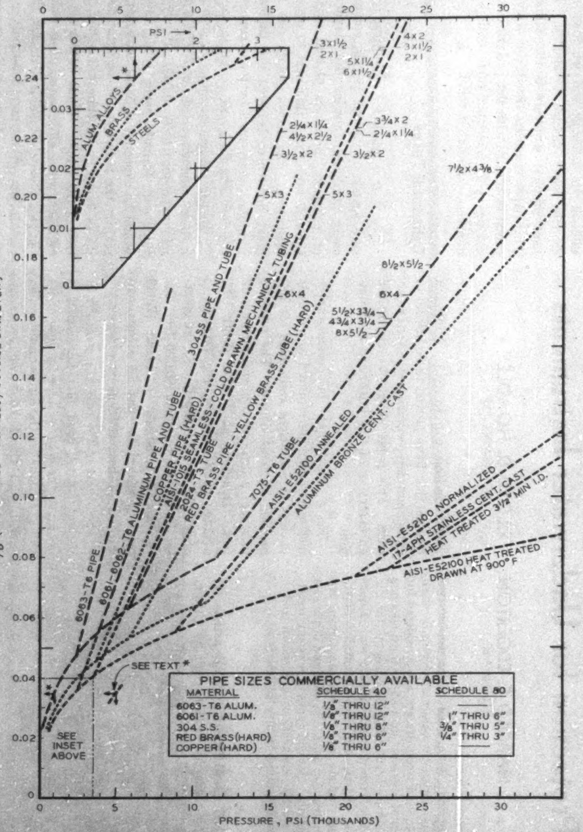
Cylindrical Tubes

Fig. 1 is a graph for cylindrical pressure vessels of various metals, commercially available, showing the maximum or collapsing pressures plotted against the t/D ratio (wall thickness to outside diameter). The figure points out the great differences between the thin-wall formulas, represented by the exponential sections, and thick-wall formulas, shown by the straight-line sections. An example is shown assuming a tube of *7075-T6* aluminum having a t/D of 0.035. If only the thick-wall formula were used, failure would be expected at 5,000 psi when, in reality, the failure would occur at 1,000 psi as shown by the thin-wall curved section of the graph. These curves are for cylinders whose effective length is greater than six times the outside diameter, since this size range is the most commonly used, and were obtained by combining formulas of Timoshenko, Carmen, and Love[2] and

---

Paper 60-662, recommended by the AIEE Special Instruments and Auxiliary Apparatus Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE North Eastern District Meeting, Providence, R. I., May 2–4, 1960. Manuscript submitted January 25, 1960; made available for printing August 10, 1960.

Harold E. Edgerton is with Massachusetts Institute of Technology, Cambridge, Mass., and Lloyd D. Hoadley is with Woods Hole Oceanographic Institution, Woods Hole, Mass.



PIPE SIZES COMMERCIALLY AVAILABLE

MATERIAL	SCHEDULE 40	SCHEDULE 80
6063-T6 ALUM.	⅛" THRU 12"	—
6061-T6 ALUM.	⅛" THRU 12"	1" THRU 6"
304 S.S.	⅛" THRU 8"	⅜" THRU 5"
RED BRASS (HARD)	⅛" THRU 6"	¼" THRU 3"
COPPER (HARD)	⅛" THRU 6"	—

Fig. 1 (left). Collapsing pressure of long cylinders as function of wall thickness to outer diameter, t/D versus psi; points showing some representative sizes of aluminum and steel tubing, available commercially

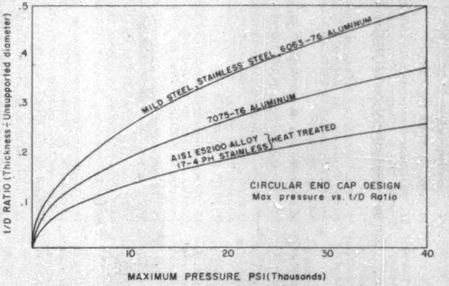


Fig. 2. Collapsing pressure for circular end-cap design as function of thickness to unsupported diameter

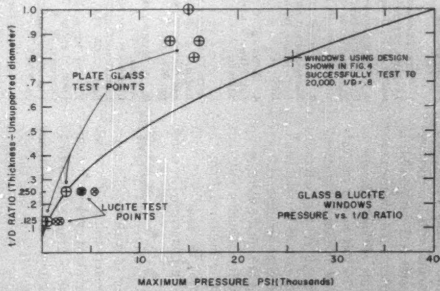


Fig. 3. Collapsing pressure of glass and Lucite windows as function of unsupported diameter

Vol
Pt.
196

further modified by experimental data from both laboratory and at sea. Slight variations from these curves may be expected since values used are averages. No allowances have been made for end-closure stresses since these are relatively small on a long tube.

Fig. 2 is a graph showing the t/D ratio (thickness/unsupported diameter) versus maximum pressure for various metallic end caps. These curves, of course, assume sufficient material to prevent shear and crushing failures. Fig. 3 gives t/D versus pressure for glass and Lucite[*] windows, showing some of the test points. From this it can be seen that a 1-inch-thick window with a 4-inch diameter ($t/D=0.25$) will break at about 2,500 psi, or approximately 1 mile. In order to use a 1-inch-thick window for greater depths, a heat-treated stainless-steel insert or ring with a $1^1/_4$-inch effective diameter is used ($t/D=0.80$, maximum $P=25{,}000$ theoretical); see Fig. 4. This design shows the effect of undercutting of the outside of the steel backup plate. This is necessary to prevent the deflection of the steel plate from affecting the glass window. Without undercutting or extremely heavy backup plate ($t/D=1$ to 1.5), failures below 50% of expected are experienced consistently.

Although regular plate glass, which is almost optically flat, has a slight greenish tint and is not the strongest window material available, it has proved to be most satsifactory. Herculite,[†] a hardened plate glass, has strengths as much as four and five times greater than that of regular plate glass. However, the slight waviness that may be introduced in the hardening process makes the seating difficult and gives highly unpredictable pressure results. These results vary from the calculated values for Herculite to values lower than for ordinary plate glass.

Fused ground quartz has been used to gain optical qualities in special cases, but for general work the slight transmission gain does not outweigh the cost ratio and there is no gain in strength.

The surfaces of the glass windows, especially the inside, must be without scratches since fracture will start at any blemish when the glass is under tension. An optical polish to a wavelength of light is necessary for the camera window. Flat Lucite or Plexiglas[‡] has proved very useful in the lower pressure ranges. Fig. 3 shows that the curve is conservative for Lucite design. Optical designers have indicated that their preference is glass, because Lucite is easily scratched and shows a loss in photographic quality because of internal stresses.

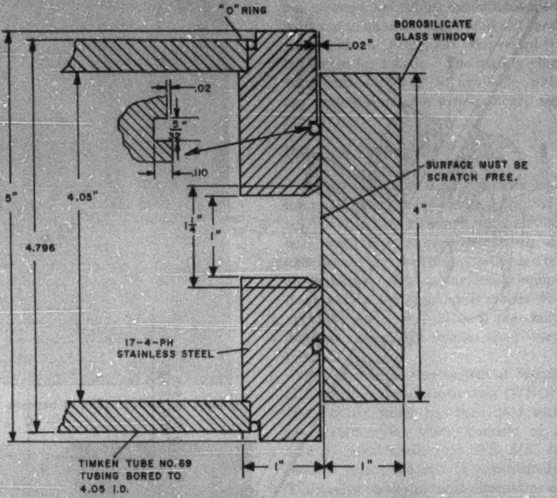

"O" RING
BOROSILICATE GLASS WINDOW
.02"
.02
3/32
.110
SURFACE MUST BE SCRATCH FREE.
5"
4.05"
4.796
1 1/4"
1"
4"
17-4-PH STAINLESS STEEL
TIMKEN TUBE NO. 69 TUBING BORED TO 4.05 I.D.
1"
1"

Fig. 4. Details of design of optical window for use in underwater camera at pressures to 20,000 psi

Fig. 5. Conax pressure gland

Seals for Electric Conductors

Several solutions for the electric wire connections have been found. One method is to use a pressure fitting on a solid wire which is insulated with plastic material. It is impossible to seal stranded wire since the water will find a path through the area between the strands. The Conax Company, Chicago, Ill., makes a standard line of stainless-steel pressure fittings that have been satisfactorily used (see Fig. 5). It has been found that the wire and insulation must be a tight fit in the insert if the gland is to be used at very high pressures. Great care must be used to tap clean perfect threads. The use of an undersized tap drill and a taper pipe-tap reamer is almost essential. Pipe-fitting cement such as Permatex[§] no. *1* (form *A* gasket; hardening) must be used and the fitting must be installed with considerably greater torques than required for ordinary use. The use of heavy-duty hexagonal socket wrenches is recommended to prevent rounding off of the hexagonal sides of the fittings. Other various modifications, such as multiple rubber inserts, etc., are occasionally employed.

Another convenient and successful solution to the lead problem (to about 10,000 psi) is the plug (Fig. 6) made by the Joy Company, Pittsburgh, Pa. One watertight separable plug, which gives three circuits, is the Joy-type *8107*. This plug can be attached from the outside with the three wires going through three holes in the metal. O rings and clamps complete the seal on the outside with the heavy rubber body serving to withstand the pressure that tries to extrude the three wires into the cavity. This seal is adequate to 10,000 psi at 20 C (degrees centigrade). The wires break due to rubber flow when used at about 20,000 psi for an extended time at 20 C. Perhaps the operation at lower tempera-

---

[*] Trade-mark, E. I. du Pont de Nemours and Company, Inc., Wilmington, Del.

[†] Trade name, Pittsburgh Plate Glass Company, Pittsburgh, Pa.

[‡] Trade name, Rohm & Haas Company, Philadelphia, Pa.

[§] Trade name, Permatex Company, Inc., Kansas City, Kan.

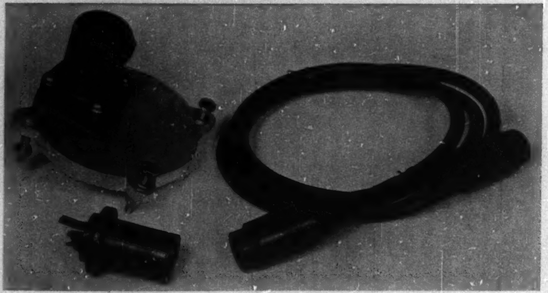


Fig. 6 (left). Watertight plug made by Joy Company, type 8107

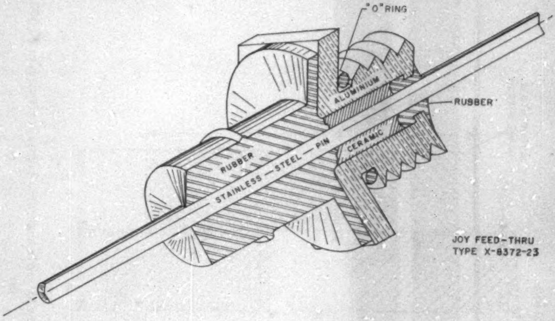
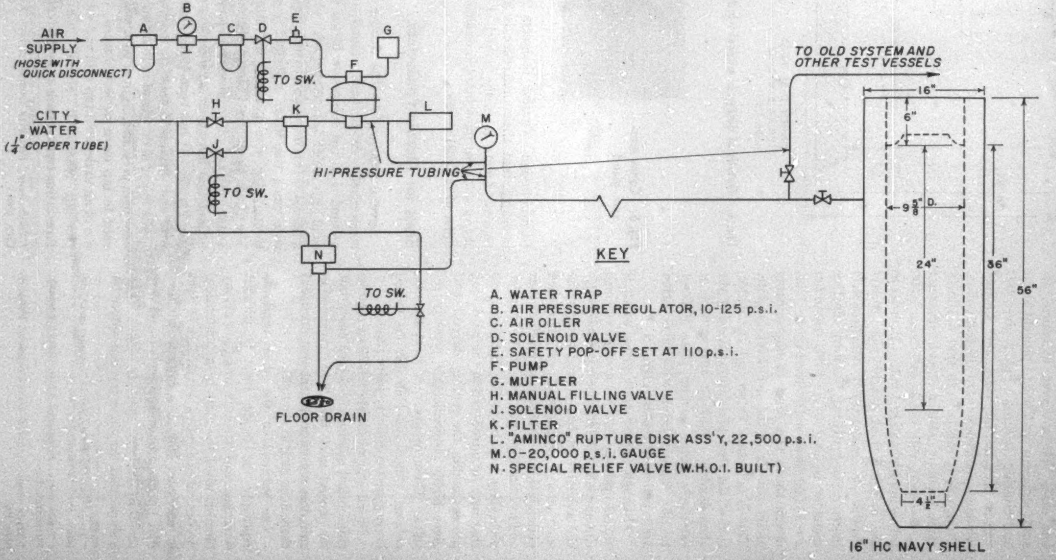

"O" RING
ALUMINIUM
RUBBER
RUBBER
STAINLESS — STEEL — PIN
CERAMIC
JOY FEED-THRU
TYPE X-8372-23

Fig. 7 (right). Sectionalized drawing of Joy feed-through electric lead which can withstand 20,000 psi


AIR SUPPLY
(HOSE WITH QUICK DISCONNECT)
CITY WATER
($\frac{1}{4}$" COPPER TUBE)
A
B
C
D
E
F
G
H
J
K
L
M
N
TO SW.
TO SW.
TO SW.
HI-PRESSURE TUBING
FLOOR DRAIN
TO OLD SYSTEM AND OTHER TEST VESSELS
16"
6"
9$\frac{5}{8}$" D.
24"
36"
56"
4$\frac{1}{2}$"
16" HC NAVY SHELL

KEY

A. WATER TRAP
B. AIR PRESSURE REGULATOR, 10-125 p.s.i.
C. AIR OILER
D. SOLENOID VALVE
E. SAFETY POP-OFF SET AT 110 p.s.i.
F. PUMP
G. MUFFLER
H. MANUAL FILLING VALVE
J. SOLENOID VALVE
K. FILTER
L. "AMINCO" RUPTURE DISK ASS'Y, 22,500 p.s.i.
M. 0-20,000 p.s.i. GAUGE
N. SPECIAL RELIEF VALVE (W.H.O.I. BUILT)

Fig. 8. Details of 20,000-psi pressure chamber which uses 16-inch U. S. Navy shell. Aminco is a trade name for the American Instrument Company, Silver Spring, Md.

Vol.
Pt. 1
196

Fig. 9. Deep-sea camera showing how chassis is arranged to utilize cylindrical housing

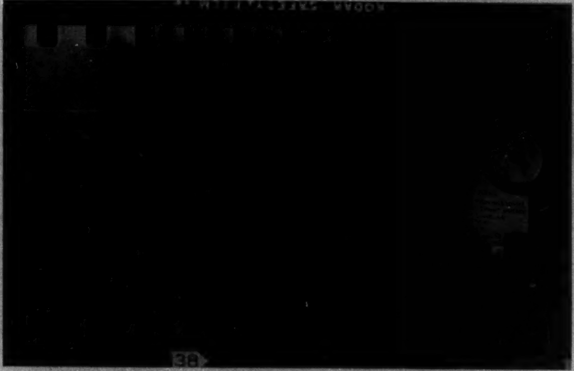


Fig. 10. Example of underwater photograph taken in pressure housing, data section photographed on each frame

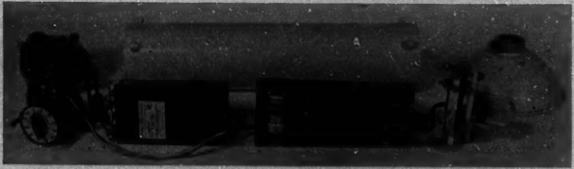


Fig. 11. Electric flash lamp and driving circuit arranged to be installed in cylindrical housing for underwater use

tures will be more favorable since the rubber hardens.

A single-pin plug, Joy-type *8372*, has been tested successfully at 14,000-psi pressure for an hour at room temperature. Failure due to shear in the rubber occurred at 20,000 psi after a sustained pressure at room temperature. At about 5 C this plug withstood 17,000 psi for 1 hour.

A newer type single-pin Joy plug no. *8372-23*, a combination of a stainless-steel pin, ceramic, and rubber insulation sealed with and to an aluminum hexagonal head screw with an O ring, has been tested to 20,000 psi for 1 hour. The mating plug with cord *x-8372-72* is available with any desired conductor arrangement; see Fig. 7.

Pressure Testing

A very important facility for developing underwater equipment is a pressure-testing chamber. Pressure tests should be applied in a routine manner to all equipment that is scheduled to be used on an expedition. In this way, faulty equipment will be found before it causes disappointments at sea. This is especially important since ship time at sea is very precious.

The pressure-testing system at Woods Hole Oceanographic Institution (WHOI) (see Fig. 8), which has been used with great success for years, consists of a 16-inch naval shell, type *HE MK 13* mods *5* (mods *0-6* can be substituted). The cavity has cylindrical dimensions of $9^5/_8$-inch diameter by 30 inches with a pointed end extending to 42 inches. The base plug is about 6 inches deep with a left-hand square thread. A standard 1/4-inch O ring is used with a brass ring insert to set the correct squeeze on the O ring. This plug weighs about 50 pounds. Another smaller hole (about $1^1/_2$ inches in diameter) at the center is closed with another left-hand thread on a screw plug with an O ring seal. This is the last plug to be placed so that the water can be observed as the shell is filled prior to the application of pressure. It is essential that no air be entrapped in the pressure system, in order to lessen the danger in case of a mechanical failure. Hard O rings of durometer 90 should be used.

Stainless-steel tubing of 1/4-inch outside diameter 0.065-inch wall and Parker* triple fittings are used to connect the hydraulic pressure pump to the shell. Oil pumps have been used, both hand and motor driven, with an oil–water interchange for pressures to 10,000 psi.

Pressures to 20,000 psi are obtained from a Sprague-Symons† pump, type *S440-200*, that operates from air pressure at 100 psi and uses water as a pressure medium.

It might be advisable to note at this point that the suitable factor of safety of four is not incorporated in this equipment, but since it is used only occasionally and the operators do not remain in the area while testing, personal safety is assured.

---

* Trade name, Union Insulating Company, Inc., Parkersburg, W. Va.

† Trade name, Symons Associates, Inc., Tiverton, R. I.

Pressures as recorded on the gage are read through an inexpensive telescope from a safe vantage point. Also, the total volume of water pumped due to the compressibility of the water and the "stretch" of the steel shell amounts to approximately 2 liters, so that even the slightest leak is noticed by the corresponding pressure drop.

In all tests it is advisable to maintain pressure for at least twice, if not four times, the expected use time to allow for plastic creep or flow that might give false confidence gained by a "pop" test.

Camera

Since conventional cameras are not easily adaptable to cylindrical shape, an in-line camera design was evolved from the standard (no. *10*) 100-foot 35-millimeter motion-picture reel. Fig. 9 shows the way in which the camera parts are lined up on a chassis that slips into the cylindrical housing. An inside diameter of 4.04 inches is required for the film spools.

A sample photograph is shown in Fig. 10. Notice that a data section has been added to each frame to record time, pressure (2,000 meters), and other information concerning the test. In this way, the essential information concerning a film always goes with the film, precluding any possibility of mismarking or mixups which often occur in processing and printing.

Electronic Flash

The elements of an electronic flash lamp are shown in Fig. 11 in a cylindrical form for insertion into a 4.04-inch-inside-diameter cylinder. The flash lamp is protected by a $1^1/_2$-inch iron pipe size Pyrex glass end cap made by the Corning Glass Works. The open end is ground flat and sealed to the metal end plate with a 1/16-inch-thick rubber gasket. It is important to use spring-lock washers on the bolts to take up the compression of the rubber, otherwise a leak may occur when the case is brought to the surface.

References

1. Photography of the Ocean Bottom, **M. Ewing, A. Vine, J. L. Worzel.** *Journal*, Optical Society of America, New York, N. Y., vol. 36, 1946, pp. 307–21.

2. Mechanical Engineers' Handbook (book), **Lionel S. Marks**, editor. McGraw-Hill Book Company, Inc., New York, N. Y., 1951.

3. Cameras and Lights for Underwater Use, **H. E. Edgerton, L. D. Hoadley.** *Journal*, Society of Motion Picture and Television Engineers, New York, N. Y., July 1955, pp. 345–50.

4. Deep-Sea Photography, **R. N. Harvey, E. R. Baylor.** *Journal of Marine Research* (*Sears Foundation*), New Haven, Conn., vol. VII, no. 1, Apr. 1948, pp. 10–16.

5. Submarine Photography with the Benthograph, **K. O. Emery.** *The Scientific Monthly*, Washington, D. C., vol. 55, no. 1, July 1952, p. 1.

6. A Wide-Angle, Underwater Camera Lens, **E. M. Thorndike.** *Journal*, Optical Society of America, vol. 40, 1950, p. 823.

7. The High-Speed Photography of Underwater Explosions, **P. M. Fye.** *Journal*, Society of Motion Picture and Television Engineers, Oct. 1950.

8. Underwater Photography (book), **H. Schenck, H. Kendall.** Cornell Maritime Press, Cambridge, Md., 1954.

9. Photographing the Sea's Dark Underworld, **H. E. Edgerton.** *National Geographic Magazine*, Washington, D. C., Apr. 1955, p. 253.

10. Deep-Sea Cameras of the Lamont Observatory, **E. M. Thorndike.** *Deep-Sea Research*, New York, N. Y., vol. 5, 1958.

11. Suspended Echo-Sounder and Camera Studies of Mid-Water Sound Scatterers, **H. R. Johnson, R. H. Backus, R. H. Hersey, D. M. Owen.** *Ibid.*, vol. 3, 1956, p. 266.

12. The Luminescence Camera, **H. E. Edgerton, L. R. Breslau.** *Journal*, Biological Photographic Association, Milwaukee, Wis., May 1958; also Woods Hole Oceanographic Institution, Woods Hole, Mass., 1958.

13. NEL Type III Deep-Sea Camera, **C. J. Shipek.** *Report no. 768*, U. S. Navy Electronics Laboratory, San Diego, Calif., Mar. 13, 1957.

14. A New Deep-Sea Camera, **A. S. Laughton.** *Deep-Sea Research*, vol. 4, 1957, pp. 120–25.

15. Aluminum Structural Design Handbook Reynolds Metals Company, Louisville, Ky., 1957

Experiments with Air Ions on the Trachea

JOHN C. BECKETT
MEMBER AIEE

ALBERT P. KRUEGER
NONMEMBER AIEE

MAN REGULARLY encounters changing environmental conditions. Some of these changes are severe involving extreme discomfort, while others are so subtle as to pass unnoticed. Air ionization is one factor of environment that is properly classified as a subtle change for most people. Few are consciously aware of any change in air ionization, yet the physiological importance of air ions to the function of the respiratory system has been recently quantitatively demonstrated.[1]

Radiation and ionization are closely allied factors of environment. Air ions or electrical charges in the air are a direct result of radiation energy affecting molecules in the air. At the earth's surface radioactivity of the soil is the principal radiation energy causing ionization. Cosmic rays are another important source, particularly as altitude increases. These ions from all sources vary in number and size depending on atmospheric conditions and the strength of radiation energy, but there are always ions present in air.

Delwiche and Stout have observed that changes in radioactivity of the atmosphere may take place rapidly, and a high value may fall to a low value within an hour or less.[2] This observation correlates closely with measurements of air ion density. Indeed, there is measurable correlation of radon gas with air ions near the earth's surface. Most of the ions in air close to the ground are the result of gamma rays from radioactive material in the soil and rock and alpha particles from radon gas emanating from porous earth. Cosmic rays are the second most important source of air ions over land. Maximal values are typically found in still air over land, and it is common to observe an ion density peak between 4:00 and 6:00 a.m. when the air is usually quiet. A noticeable drop in ion density after sunrise can be attributed to convection currents acting to dilute the accumulation of radon gas with greater masses of air. During hot weather the ion density often reaches a minimum about midafternoon.

It has been generally known for more than 150 years that air contains electricity. This was a popular subject of investigation as recently as the 1930's, but it has only been since 1950 that means for control of air ion density as an environmental factor has been given serious consideration.[3]

Air Ion Measurement

There are many designs of ion collectors for measuring air ion density, including those of Gerdian, Ebert, Zeleny, Erickson, Van de Graff, and Israel.[4–6] All of these devices are similar in principle, including both the concentric and parallel-plate-type units. Ion collectors have been constructed with two or more stages in series to measure more accurately the distribution of ions of different size. It is assumed that each ion has one unit charge, and this has been shown by Loeb[7] to be highly probable on particles of the size under consideration. In a uniform electric field a large ion with one unit charge (1.6×10^{-19} coulombs) will move slower than a small ion; thus, the ion mobility is inversely proportional to size. Simultaneous recording of the number of ions present in several mobility ranges provides a clearer picture of the size distribution of ions in air. Small ions are described as gaseous ions consisting of 3–8 molecules, depending on age and environmental conditions. The average mobility of these ions is between 1 and 2 centimeter/second/volt/cm.

Accurate measurement of ion mobility has been made in the physics laboratory with dry air.[8] Ion collectors for general use under various conditions of atmospheric pressure, temperature, and humidity provide a reasonable picture of the small ion "mobility spectrum"[9] estimated to be accurate within about 15%. Principal sources of error are air turbulence and the random position of ions as they enter the collector.

One type of ion collector that is particularly convenient for environmental studies has a first-stage ion trap for selectively removing high-mobility ions in predetermined steps calculated to provide a recording of the ion mobility distribution from the smallest ion through the intermediate range (4 to 0.002 cm/sec/volt/cm). The ion trap voltage is varied in logarithmic steps from 1/3 volt to 600 volts at intervals of 15 sec/step. See Figs. 1 and 2.

Another device for measuring ions is the target collector. In its simplest form this is a gold-leaf electroscope equipped with a target electrode. The time of discharge is inversely proportional to the rate of arrival of unipolar ions. A more convenient ionization flow meter of this type has been developed for the laboratory. See Fig. 3. An electrostatically shielded metal disk about 5 cm² in area is connected to an electrometer for direct reading of either polarity. This same principle is used to measure the air-to-earth current in studies of atmospheric electricity. Detection and accurate measurement of ion flow from an ion source in an electric field is reduced to the simplest form with this device. Rate of ion flow is reported in units of ions per cm² of surface area per second.

Ion Generation

The principle of ion generation and control is simple. Unipolar ion generation employs a means to produce both positive and negative ions and an electric field to effect ion separation. Desired ions can then be disseminated by means of the same electric field or with forced air passing over the ion source. Control implies careful monitoring of ion density, use of ion filtering and ion generation to compensate for natural variations.

It is possible with ion measuring equipment to demonstrate the many sources of air ions and the ease with which ions can be generated. High-voltage electric fields, incandescent materials, ultraviolet light, X ray, and radiation from radioactive isotopes have all been used. Where precise ion control is desired, that is, a uniform output of one or both polarities of ions adjustable to meet specified conditions, radioactive sources have been the most acceptable. Polonium 210, an alpha emitter, and tritium, a soft beta emitter, are commercially available for this purpose.

An ion generator for the control of room air is shown in Fig. 4. This compact portable device contains an ion source consisting of 50 millicuries of tritium gas sealed in zirconium deposited

---

Paper 60-101, recommended by the AIEE Electrical Techniques in Medicine and Biology Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., January 31–February 5, 1960. Manuscript submitted November 4, 1959; made available for printing January 29, 1960.

JOHN C. BECKETT, formerly with the Wesix Electric Heater Company, San Francisco, is now with the Hewlett-Packard Company, Palo Alto, Calif.; and ALBERT P. KRUEGER, M.D., is with the University of California, Berkeley, Calif.

The physiological experiments quoted herein were supported by contracts with the Atomic Energy Commission, United States Air Force (A.F. *49(638)-669*) monitored by the Air Force Office of Scientific Research of Air Research and Development Command and by grants from the National Institutes of Health (*E-2837*) and the Committee on Research of the University of California.

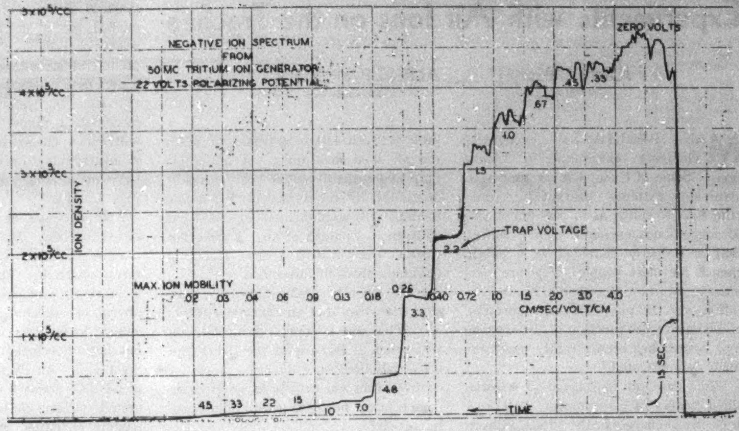


Fig. 1. Recording of negative ion mobility distribution using a parallel-plate-type ion collector equipped with an ion trap selectively to remove ions

on stainless steel foil. An electric potential difference of approximately 2,000 volts direct current is provided to separate the positive and negative ions before they are lost through recombination in the dense ion plasma surrounding the foil.[10] Tritium is a low-energy 0.0185-mev (meg-electron-volt) beta source and produces ions as a result of beta particles ionizing air through a range of about 1 cm at atmospheric pressure. The desired ions are directed by the electric field into a moving airstream created by a small fan located behind the ionizing element. The quantity of tritium used in these experiments produces not less than 10^9 ions/second in air for distribution to the work area. Siksna has shown that the maximum utilization of available ions for room air is achieved using forced air.[11] In tissue studies forced air complicates the experiment and an electric field alone is preferred for laboratory research. In the physiological studies reported here the tritium ionizing element was positioned within the animal or tissue enclosures and circulating air at low velocity was used only in long term animal studies extending for a period of several days or weeks. Typical ion distribution from a tritium ionizing element operating in ordinary air is shown in Fig. 2. Ordinary air is defined as normal air as found in rural areas free of industrial or automobile exhausts, or tobacco smoke. The small ion density is reduced if there are airborne contaminants, particularly those of submicroscopic size.[12]

Electrostatic ion generators using 5,000 volts direct-current on fine tungsten wires have been successfully developed but are

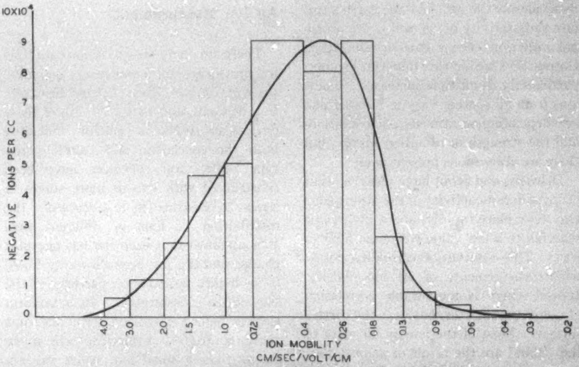


Fig. 2. Plot of data recorded in Fig. 1 showing "mobility spectrum"

Note: In the record in Figs. 1 and 2 of ions from a tritium source ion generator, small ions are trapped in a field less than 15 volts/cm within a travel distance of 7 cm at an air velocity of 75 cm/sec. Record is from a 50-millicurie tritium negative ion generator of the type illustrated in Fig. 4. The shape of this curve is approximately the same as found in nature under similar conditions

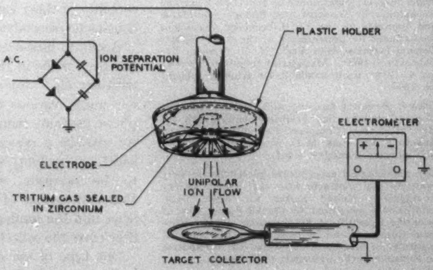


Fig. 3. The target collector may be used to measure total unipolar ion current from an ion source, or to measure ion flow per unit area when positioned close to the tissue surface

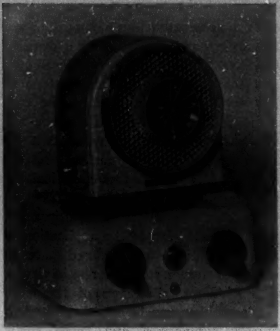


Fig. 4. Negative ion generator using a sealed tritium element containing 50 millicuries of tritium and an ion separating potential of 2,000 volts combined with a small circulating fan

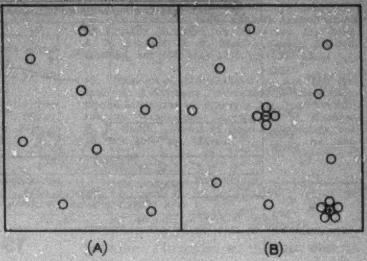


Fig. 5 (right). A small ion in air consists of a single charge attached to one molecule but associated with several to form a molecule complex. Take away the charge and this complex loses all identity

A—Molecules in air move independently and at random in absence of an electric charge

B—The essence of a small ion is the electric charge. Several molecules (3–8) become associated with a single charge, positive or negative

more susceptible to variations in ion output. Ultraviolet and thermionic sources have not been satisfactory for scientific studies because of the rapid deterioration of ion output. Ozone is often a by-product of high-voltage and ultraviolet-type ion generators and the undesirable nature of ozone makes it mandatory to use such generators with caution.

The Influence of Air Ions on the Windpipe (Trachea)

It is not at all obvious that small air ions should exert any kind of effect on living systems. The potential energy involved is extremely small, though the total energy contributed over a long period of time may be significant. Ions move relatively slowly, e.g., 1 to 2 cm/sec in a field of 1 volt/cm, and consequently cannot be expected to inflict physical damage on impact. Further, they affect only about one molecule in 5 trillion.[13] Nevertheless, these ions are easily detected by means of simple equipment and there can be no doubt about their ability to influence the surface-clearing efficiency of the mammalian windpipe (trachea) as measured by the rate of mucus flow and the beat rate of microscopic hair-like protrusions called cilia. It is the small ions that dominate in clean air and are of particular interest. The essence of the small ion is its charge; take away the charge and the gaseous molecule complex associated with the charge loses all identity. See Fig. 5. This is not the case with charged particulate matter.[14]

First experiments were performed with tracheal strips of the rat, rabbit, and guinea pig, kept in a moist chamber and arranged in such fashion that their surfaces could be viewed through a microscope equipped with an eye-piece micrometer. This made it possible to measure the rate of mucus flow by timing the progress of air bubbles in the mucus layer. Similarly, a strobotachometer providing an intense light source flickering at variable rates permitted determination of the rate of ciliary beat with an accuracy of ±50 beats/minute. See Fig. 6.

Air ions were generated within the chamber by soft beta radiation derived from tritium contained in a sealed foil; a reversible rectifying circuit allowed selection of positive or negative ions at will. The number of ions impinging on each cm^2 of surface/sec was measured with a target collector located in the same relative position as the tracheal strips and connected to a Beckman micromicroammeter.

Moderate densities ($1x10^7/cm^2/sec$) of positive ions acting for 20–30 minutes produced striking effects on the functioning of the trachea. Measurement of ions received on the tissue was accomplished immediately after observing ciliary beat rate using the device shown in Fig. 3 in place of the tracheal strips. Ciliary activity dropped by several hundred beats per minute or frequently stopped altogether, while the rate of mucus flow diminished. Positive ions contracted the posterior wall of the windpipe (trachea), dried the mucosal surface, and induced in the cilia a singular susceptibility to tissue damage (trauma). Minor mechanical stimuli that ordinarily had no effect now resulted in complete cessation of ciliary action.[15]

Negative ions accelerated ciliary activity and mucus flow above normal values and in general reversed the inhibitory effects of positive ions.

It was found that similar results could be obtained in the living animal by administering the ions to the undisturbed lining of the windpipe (trachea) through a tracheotomy aperture.[1] In addition to the functional changes noted in the isolated tracheal strip, the living animal exposed to positive ions displayed a pronounced increase in the susceptibility of the mucosal vessels to damage (trauma). The slightest pressure with a probe, which normally results in no notable vascular reaction, induced the prompt formation of a bruise (ecchymotic disc).

The next set of experiments was the exposure of unoperated, ambulatory animals to unipolar ionized atmospheres. Providing somewhat longer periods of exposure were used, they developed the identical alterations in the functional

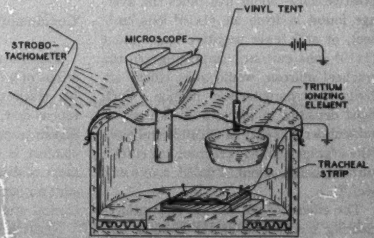


Fig. 6. Experimental arrangement for exposure of extirpated tracheal strip to air ions from a tritium source

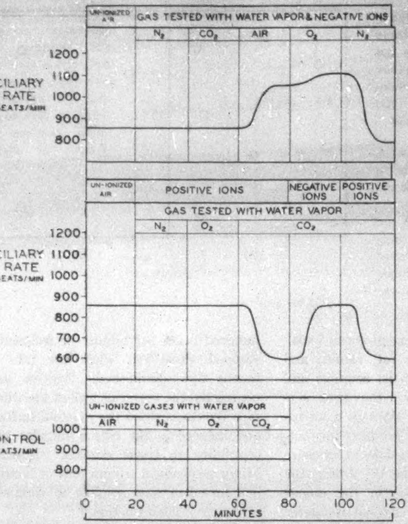


Fig. 7. Ciliary activity of anesthetized rabbits in un-ionized air and in three atmospheres, consisting of N_2, CO_2, and O_2

activity of the trachea which has already been described.

Curiosity leads one to wonder: 1. What constituents of ionized air are responsible for the observed alterations in function? 2. Are these responses to air ions essentially a laboratory phenomenon or do they occur in nature?

Experimental evidence established that negatively charged oxygen is the physiological mediator of negatively ionized air.[16] Experiments on tissue showed that only during exposure to oxygen would the ciliary beat rate increase. Nitrogen, carbon dioxide, and water vapor negative ions had no effect. Similarly, positively charged carbon dioxide is the active component of positively ionized air. Only in the presence of measurable amounts of carbon dioxide would positive charges decrease the ciliary beat rate of trachea strips. See Fig. 7. Further work with the tracheal strip and the tracheotomized animal clearly showed that a very few positive ions—about five times the average found in fresh air (1×10^4 ions/cm²/sec), or still fewer negative ions—about (1×10^3 ions/cm²/sec) impinging on the tracheal mucosa, would alter ciliary activity in characteristic fashion. Such light densities of air ions are known to exist under a variety of natural conditions (the average quantity reaching the earth's surface being about 2.5×10^3 ions/cm²/sec).[17]

The effects of air ions on the surface-clearing mechanism of the trachea and larger bronchi, as seen in the laboratory, are not limited to one particular species but apply to monkeys and man as well as to the usual experimental animals.[18] The inhibition of the clearing function by positive ions and its acceleration by negative ions appear to be entirely independent of the central nervous and circulatory systems and are triggered by relatively few ions impinging on the mucosal surface.

Recent studies have proved that negative air ions stimulate the activity of cytochrome oxidase thereby accelerating the process of reoxidation of reduced cytochrome *c*.[19] Such action on a respiratory enzyme normally found in tissues may account for the effects of negative ions on ciliary activity and perhaps on other functions as well. Positive ions exerted no influence on this particular enzyme system and appear at this time to have quite a different effect, unrelated to the negative ion enzyme action.[20]

Conclusion

Though an individual may not be conscious of a change in ciliary rate in his or her trachea, or aware of an increase in mucus flow, the importance of these findings in terms of environmental health and comfort is obvious. The specific treatment of disease, on the other hand, must wait for completion of carefully controlled clinical studies.

Negative air ions have been shown to be beneficial to respiration, but this does not mean that negative ions alone will cure a respiratory ailment or relieve a diseased condition. There is no evidence to suggest that positive ions cause asthma or hay fever; the most that can be concluded is that positive ions enhance susceptibility to trauma, may account for some mucosal irritation and may set the stage for infection if other conditions are also present.

Electrical engineers will appreciate their responsibility in the application of electrical equipment in confined spaces to avoid positive ion space charges in view of these findings. Extremes of the engineer's problem can be found on modern nuclear submarines and space ships. Closed atmospheric systems need careful study of air ions as a part of providing the most favorable environment. At the scientific level, the biophysicists and physiologists have a new challenge in the study of electrical effects on cells and tissue. Previously it was not suspected that such small quantities of electricity could cause any measurable response. Improved measuring devices and techniques have helped to make this possible.

References

1. Effects of Air Ions on the Living Mammalian Trachea, A. P. Krueger, R. F. Smith. *Journal of General Physiology*, New York, N. Y., vol. 42, no. 1, Sept. 1958, pp. 69–82.

2. Diurnal and Other Fluctuations in Atmospheric Radioactivity of Natural Origin, C. C. Delwiche, P. R. Stout. *Bulletin*, American Meteorological Society, Boston, Mass., vol. 40, June 1959.

3. Air Ionization as an Environment Factor, J. C. Beckett. *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 73, Sept. 1954, pp. 161–65.

4. Atmospheric Electricity, J. A. Chalmers. Pergamon Press, Inc., New York, N. Y., 1957, pp. 44–54.

5. Terrestrial Magnetism and Electricity (book), J. A. Fleming. Dover Publications, Inc., New York, N. Y., 1949, p. 252.

6. Air Ionization and Its Hygienic Significance (book, in Russian), A. A. Minkh. Medgiz, Moscow, U.S.S.R., 1958, chap. IV.

7. Fundamental Processes of Electrical Discharge in Gas (book), L. B. Loeb. John Wiley & Sons, Inc., New York, N. Y., 1947, p. 49.

8. *Ibid.*, p. 32.

9. *Ibid.*, p. 41.

10. Production of Unipolar Air with Radium Isotopes, Thomas. L. Martin, Jr. *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 72, 1953 (Jan. 1954 section), pp. 771–76.

11. Air Ions Produced by a Tritium-Ion Generator, Reinhards, Siksna, Robert Lindsay. *Arkiv for Geofysik*, Almquist & Wiksell, Stockholm, Sweden, vol. 3, no. 8, 1959, pp. 123–39.

12. Dynamics of Fresh Air, J. C. Beckett. *ASHRAE Journal*, American Society of Heating, Refrigerating, and Air Conditioning Engineers, New York, N. Y., vol. 1, no. 5, May 1959, pp. 47–51.

13. The Physiological Significance of Positive and Negative Ionization of the Atmosphere, A. P. Krueger, R. F. Smith. *Journal*, The Royal Society for The Promotion of Health, London, England, vol. 79, no. 5, Sept.–Oct. 1958, p. 644.

14. Atmospheric Electricity, J. A. Chalmers. Pergamon Press, Inc., 1957, p. 57.

15. Effects of Air Ions on Isolated Rabbit Trachea, A. P. Krueger, R. F. Smith. *Proceedings*, Society of Experimental Biology and Medicine, New York, N. Y., vol. 96, 1957, pp. 807-09.

16. Further Studies of Gaseous Ion Action on Trachea, A. P. Krueger, R. F. Smith, G. J. Hildebrand, C. E. Meyers. *Proceedings*, Society of Experimental Biology and Medicine, vol. 102, 1959, pp. 355-57.

17. Parameters of Gaseous Ion Effects on the Mammalian Trachea, A. P. Krueger, R. F. Smith. *Journal of General Physiology*, vol. 42, no. 5, May 1959, pp. 959-69.

18. Effects of Air Ions on Trachea of Primates, A. P. Krueger, R. F. Smith, J. W. Millar. *Proceedings*, Society of Experimental Biology and Medicine, vol. 101, 1959, pp. 506-07.

19. An Enzymatic Basis for the Acceleration of Ciliary Activity by Negative Air Ions, A. P. Krueger, R. F. Smith. *Nature*, London, England, vol. 183, May 1959, pp. 1332-33.

20. Biological Mechanisms of Air Ion Action—I. 5-Hydroxy Tryptamine as the Endogenous Mediator of Positive Air Ion Effects on the Mammalian Trachea, A. P. Krueger, R. F. Smith, *Journal of General Physiology*, vol. 43, Jan. 1960, pp. 533-40.

Power Semiconductor Ratings Under Transient and Intermittent Loads

F. W. GUTZWILLER
MEMBER AIEE

T. P. SYLVAN
NONMEMBER AIEE

METHODS for rating semiconductor power rectifiers and controlled rectifiers under steady-state conditions are well established and accepted. These techniques assume uniform temperature across the entire semiconductor junction area and limit this temperature under different load and ambient conditions to a predetermined level consistent with good performance and high reliability.[1,2] Supplementary methods for determining junction temperature under transient and intermittent loads are necessary for such applications as motor-starting and lamp-dimming controls,[3] and for calculations involving other types of overload duty on semiconductor equipment. This paper discusses methods of handling these intermittent duty types of applications. Examples illustrate suggested techniques for determining the peak junction temperature under various kinds of intermittent loading.

Semiconductor Thermal Circuit

Compared with other electric components such as transformers and motors, semiconductors have very low thermal capacity, particularly in the immediate vicinity of the junction, which is the source of a major part of the heat losses. As a result, semiconductors heat up quickly upon application of load. Yet, for very short overloads this relatively low thermal capacity may be significant in arresting the rapid rise of junction temperature. In addition, the heat sink to which the semiconductor is attached may have a thermal time constant of many minutes. Both of these effects can be used to good advantage in securing attractive intermittent ratings, sometimes well in excess of the published continuous ratings for a device.

Fig. 1(A) illustrates the cross section of a typical silicon semiconductor device, in this case a diode rectifier. The main elements of the device are the *P* type and *N* type of silicon which form the active junction at their interface, suitable back-

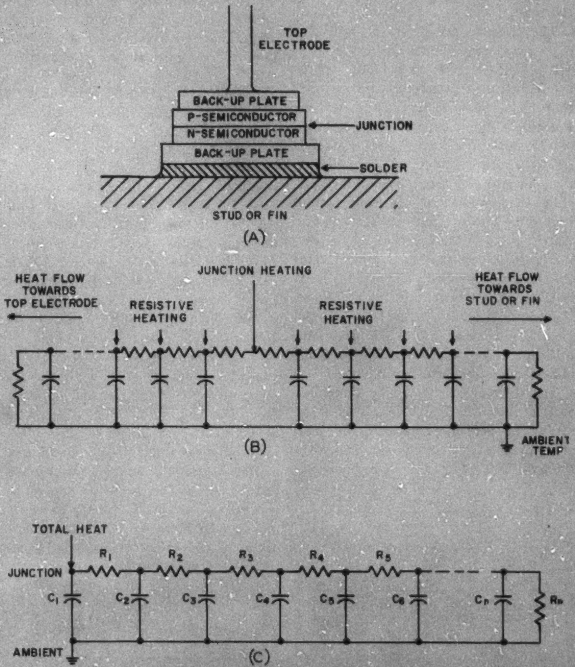


Fig. 1. A—Section through typical silicon rectifier. B—Representative thermal circuit of silicon rectifier. C—Equivalent thermal circuit (heat input at single point)

Paper 60-69, recommended by the AIEE Semiconductor Rectifiers Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., January 31–February 5, 1960, and re-presented for discussion only, at the AIEE Fall General Meeting, Chicago, Ill., October 9–14, 1960. Manuscript submitted October 30, 1959; made available for printing July 15, 1960.

F. W. Gutzwiller and T. P. Sylvan are with the General Electric Company, Auburn, N. Y., and Syracuse, N. Y., respectively.

The authors are indebted to William Warburton of the General Electric Company Low Voltage Switchgear Department for first suggesting, in a 1957 report, the concept of transient thermal resistance.

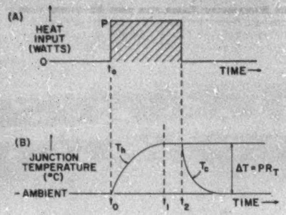


Fig. 2. Response of junction temperature to pulse of heat input

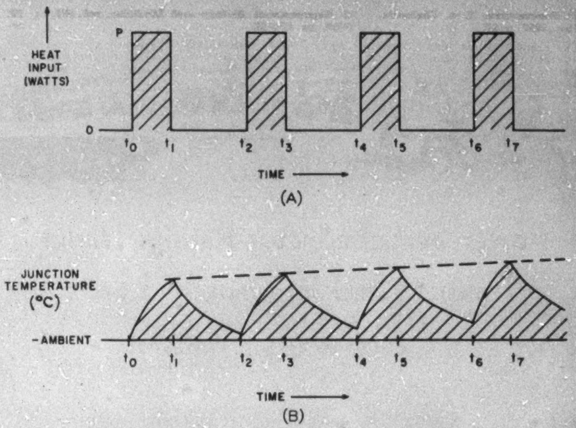


Fig. 4. Transient heating of a semiconductor under intermittent load

up plates for matching the coefficient of expansion of the silicon, and top and bottom electrodes to act both as electric connections and a path for cooling the cell. At normal load current levels, the greatest part of the total heat generation occurs at the junction with relatively insignificant I^2R (current resistance) losses occurring elsewhere in the semiconductor material, backup plates, electrodes, and the joints between these elements. At higher currents, the resistive losses become a more significant part of the total.

The thermal circuit of the semiconductor can be represented as shown in Fig. 1(B). The electrical symbol for a capacitor represents thermal capacitance, and that for a resistor, thermal resistance. For planar junctions two main networks of thermal resistance and capacity emanate from the junction to the ambient temperature: one through the materials above the junction and thence out the top lead to the ambient, the other through the materials below the junction and thence to the stud, case, fin, and ambient. In the thermal circuit of Fig. 1(B), heat is shown being introduced at the junction, as indicated by the long arrows. Heat is also generated by I^2R losses in other parts of the thermal network, as indicated by the shorter arrow.

The thermal circuit can be simplified to that shown in Fig. 1(C). This is an equivalent network emanating in only one direction from the junction and with the total heat losses of the semiconductor being introduced at the junction(s) only, instead of at various other points throughout the semiconductor device. This latter simplification is valid for current amplitudes at which I^2R losses are small in comparison with the junction losses.

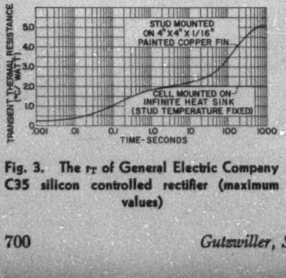


Fig. 3. The r_T of General Electric Company C35 silicon controlled rectifier (maximum values)

Junction Temperature Response to Sudden Application of Load

When a step pulse of heating power P is introduced at the junction of the semiconductor (and of the thermal circuit) as shown in Fig. 2(A), the junction temperature will rise at a rate dependent upon the response of the thermal network. This is represented by the curve T_h in Fig. 2(B). After some sufficiently long time t_1, the junction temperature will stabilize at a point $\Delta T = PR_T$ above the ambient temperature. (ΔT = temperature rise in C (degrees centigrade), P is power loss in watts, and R_T is the thermal resistance between junction and ambient). R_T is the sum of R_1 through R_n in the equivalent thermal circuit of Fig. 1(C).

If the power input pulse is terminated at time t_2 after the junction temperature has stabilized, the junction temperature will return to ambient along the locus indicated by T_c in Fig. 2(B). It is shown in the Appendix that curves T_h and T_c are conjugates of one another, i.e.,

$$T_c = \Delta T - T_h = PR_T - T_h \qquad (1)$$

Transient Thermal Resistance

By dividing the instantaneous temperature rise of curve T_h in Fig. 2(B) by the power P causing the rise, the dimensions of the ordinate can be converted from C to C/watt (per watt). This latter set of dimensions is that of thermal impedance, or, as henceforth arbitrarily termed, the transient thermal resistance r_T. Fig. 3 shows a plot of r_T for a typical medium-power semiconductor device.

The r_T for a given device can be secured by monitoring junction temperature at the end of a well-defined power pulse or after a known steady-state load has been removed. Junction temperature is measured by use of one of the temperature-sensitive junction characteristics such as reverse saturation current or forward voltage drop at low currents. Conversion of heating data to cooling data, or vice versa, can be accomplished through the use of equation 1.

To use the r_T curve with confidence in equipment designs, the curve must represent the highest values of thermal resistance for each time interval on the curve that can be expected from the manufacturing distribution of the product. An additional slight safety factor is required to provide conservative application under repetitive pulse types of load.

The r_T curve approaches asymptotic values at both the long-time and short-time extremes. For very long-time intervals the r_T approaches the steady-state thermal resistance R_T. At very short time intervals the curve approaches a limiting value which is different from zero. This asymptotic value of thermal resistance implies a peak power rating for the device so that use of the curve in the region of very short time intervals results in conservative designs. For

Vol.
Pt. 1
196

maximum utilization of semiconductor devices in the microsecond region, additional factors must be considered and other methods of rating and life testing used.

Response to Single-Load Pulse

The r_T curve forms the basis for several different types of intermittent duty calculations. The simplest is junction temperature response to a single step function of power input, as expressed in Fig. 2. Provided the junction is initially at the ambient temperature, the junction temperature rise during heating T_h can be expressed as

$$T_h = Pr_T \qquad (2)$$

where r_T is at time t (C/watt). The cooling curve T_c can be expressed as

$$T_c = P(R_T - r_T) \qquad (3)$$

where R_T = steady-state value of thermal resistance. Equation 3 is valid only if the junction has reached its steady-state temperature rise before the heat input is removed.

For example, if 100 watts are suddenly applied to the device whose r_T is defined by the infinite heat sink curve in Fig. 3, the junction temperature will rise $Pr_T = 100 \times 0.36 = 36$ C in 0.01 sec (second) and $100 \times 1.0 = 100$ C in 0.1 sec. If allowed to stabilize, the junction temperature will eventually rise $100 \times 2.0 = 200$ C above an infinite heat sink. If the power input is removed after steady-state conditions have been reached, the junction rise above the heat sink will return to $100\,(2.0 - 0.36) = 164$ C in 0.01 sec, and to $100\,(2.0 - 1.0) = 100$ C in 0.1 sec. It will return to the heat sink temperature in approximately 10 sec.

This technique is useful in calculating the maximum load-carrying capabilities of a device when subjected to single load pulses with sufficient time between pulses to allow cooling back to the initial conditions. For example, if the maximum junction temperature rating of the aforementioned device is 200 C and it is desired to determine the maximum power dissipation for 0.5 sec when the device is mounted to an infinite heat sink at 25 C, one can proceed as follows: r_T at 0.5 sec from Fig. 3 = 1.62 C/watt. Allowable temperature rise = 200 − 25 = 175 C. Maximum heating pulse = 175 C/watt/1.62 C/watt = 108 watts.

Response to Repeated Pulses

When heating pulses follow one another so closely that the junction cannot cool down to its initial temperature each cycle, the method of determining peak junction temperature becomes somewhat more complex. Fig. 4 illustrates the case where the heating pulse is not long enough for the junction to reach its steady-state rise and the OFF interval is too short for complete cooling. As a result, the junction temperature peak rises on successive cycles until, after a sufficient length of time when the temperature rise each cycle becomes the same as the temperature drop, successive peak temperatures are identical. This is the steady-state condition for that input.

Superposition: While several approximation approaches to this type of problem can be proposed, superposition[4] offers a rigorous numerical method of using the r_T curve to secure a solution, even though the steady-state conditions for equations 2 and 3 are not met. Superposition tests the response of a network to any input function by replacing the input with an equivalent series of superimposed positive and negative step functions. Each step function must start from zero and continue to infinity, to meet the steady-state requirements of equations 2 and 3.

For example, Fig. 5(A) illustrates a typical train of heating pulses. The amplitude of each successive pulse changes. Fig. 5(B) shows the equivalent heat input attained by superimposing positive and negative step functions of power. At time t_0 an infinitely long step function of heat with amplitude $+P_0$ is applied to the junction. To conclude the actual heating pulse at time t_1, an infinitely long negative step function of heat with amplitude $-P_0$ is applied to the junction. This neutralizes the initial

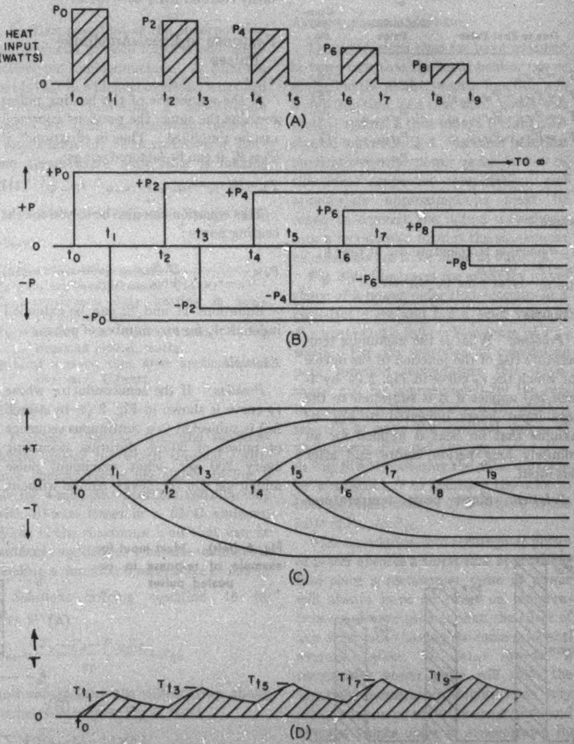


Fig. 5. Use of superposition to determine peak junction temperature. A—Heat input. B—Equivalent heat input by superposition. C—Junction temperature response to individual heat inputs in B. D—Actual junction temperature

step function from this point out to infinity. At time t_3, a positive step function P_2 is applied to initiate the second pulse. This is terminated at t_4 by a negative step function $-P_2$. The succeeding square-top pulses with varying amplitudes may be represented in similar manner.

By using the r_T curve as shown in the previous section, the temperature response of the junction may be plotted for each of the various step functions of power as shown in Fig. 5(C). Fig. 5(D) indicates the actual junction temperature variation which is simply the algebraic sum of the individual temperature response curves of Fig. 5(C).

The maximum and minimum temperatures for each cycle can be expressed in equation form, as follows:

TEMPERATURE RISE

Due te First Pulse	Due to Second Pulse	Equation No.
$T_{t_1}=P_0 r_{t_1}$		(4)
$T_{t_2}=P_0 r_{t_2}-P_0 r_{(t_2-t_1)}$		(5)
$T_{t_3}=P_0 r_{t_3}-P_0 r_{(t_3-t_1)}$	$+P_2 r_{(t_3-t_2)}$	(6)
$T_{t_4}=P_0 r_{t_4}-P_0 r_{(t_4-t_1)}$	$+P_2 r_{(t_4-t_2)}$	(7)
	$-P_2 r_{(t_4-t_3)}$	
$\vdots$		

where

T_{t_1} = junction temperature rise at time t_1
$r_{t_1}=r_T$ at time t_1
$r_{(t_2-t_1)}=r_T$ at time (t_2-t_1), etc.

Example

Problem: What is the maximum temperature rise of the junction of the device for which the r_T curve in Fig. 3 (4- by 4-inch fin) applies if it is subjected to the heat input losses indicated in Fig. 6? Assume that no heat is applied for an infinitely long period before and after this input.

Solution: Since peak temperatures will occur at the end of individual heat input pulses, we need only check the temperature at points t_1, t_3, and t_5.

$$T_{t_1}=P_0 r_{t_1}=60 \text{ watts}\times 2.19 \text{ C/watt}=131 \text{ C} \tag{8}$$

$$\begin{aligned} T_{t_3}&=P_0 r_{t_3}-P_0 r_{(t_3-t_1)}+P_2 r_{(t_3-t_2)}\\ &=60\times 2.65-60\times 2.49+50\times 2.19\\ &=159-149+110=120 \text{ C} \end{aligned} \tag{9}$$

$$\begin{aligned} T_{t_5}&=P_0 r_{t_5}-P_0 r_{(t_5-t_1)}+P_2 r_{(t_5-t_2)}-\\ &\qquad P_2 r_{(t_5-t_3)}+P_4 r_{(t_5-t_4)}\\ &=60\times 3.05-60\times 2.9+50\times 2.65-\\ &\qquad 50\times 2.49+40\times 2.19\\ &=183-174+133-124+88=106 \text{ C} \end{aligned} \tag{10}$$

It is evident that the highest temperature rise occurs at the end of the first pulse in this example. The actual junction temperature can be calculated by adding this rise to the ambient temperature. For instance, in a 50 C ambient, the junction would have momentarily reached 131+50=181 C.

Response to Equal-Amplitude Pulses

If the amplitude of the heating pulses remains the same, the previous approach can be simplified. Thus in equation 7, if $P_0=P_2$, it can be factored out as:

$$T_{t_4}=P_0(r_{t_4}-r_{(t_4-t_1)}+r_{(t_4-t_2)}-r_{(t_4-t_3)}) \tag{11}$$

This equation can also be solved for the heating power:

$$P_0=\frac{T_{t_4}}{r_{t_4}-r_{(t_4-t_1)}+r_{(t_4-t_2)}-r_{(t_4-t_3)}} \tag{12}$$

Equations 11 and 12 can be extended indefinitely for any number of pulses.

Example

Problem: If the semiconductor whose r_T curve is shown in Fig. 3 (4- by 4-inch fin) is subjected to a continuous sequence of pulses of 50-sec duration occurring every 200 sec, what maximum pulse amplitude can the device handle without exceeding a 200 C junction temperature if the ambient is 40 C?

Solution: The peak temperature rise allowable is 200 C−40 C=160 C. This is the numerator in equation 12. One can find P_0 if the denominator of this equation is available. Minimum P_0 will occur when the denominator is at a maximum. By numerical means, the series composing the denominator can be carried out to the point where it no longer increases in its total value. This will occur at approximately 1,000 sec where steady-state thermal resistance R_T is reached. At this point, the cooling terms will be equal to, and will cancel out, additional heating terms.

The following numerical solution for the denominator of equation 12 starts from the last terms and works up to the initial terms, for convenience in tabulating and solving:

Transient Thermal Resistance	C/Watt
Heating Terms	
r_{50}	+ 2.77
r_{250}	+ 4.36
r_{450}	+ 4.85
r_{650}	+ 5.04
r_{850}	+ 5.09
$r_{1,050}$	+ 5.10
Total	+27.21
Cooling Terms	
r_{200}	− 4.15
r_{400}	− 4.75
r_{600}	− 5.01
r_{800}	− 5.08
$r_{1,000}$	− 5.095
Total	−24.085

Continuing from the tabulation:

Total effective thermal resistance= +27.21 −24.085=3.125 C/watt.

From equation 12, maximum P_0=160 C/(3.125 C/watt)=51.2 watts.

Steady-State Solution: For trains of power pulses in which the period of a pulse is much shorter than the time for r_T to stabilize at R_T, the foregoing method

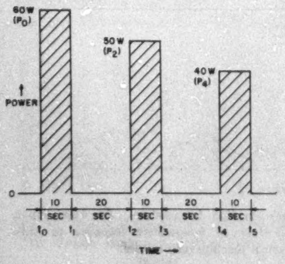


Fig. 6 (left). Heat input for example of response to repeated pulses

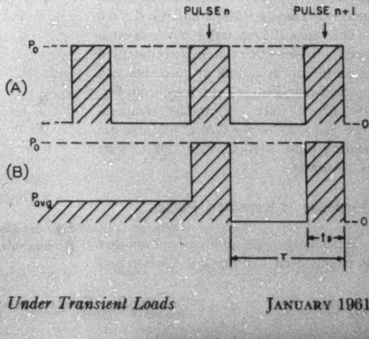


Fig. 7 (right). Heat input for steady-state pulse conditions. A—Actual. B—Assumed

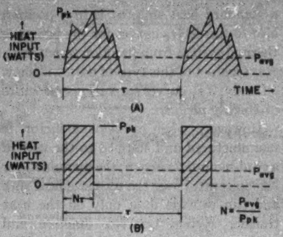


Fig. 8. Approximating an irregularly shaped power pulse by a rectangular wavesoape. A—Actual heat input. B—Equivalent heat input

of calculation may become too laborious, requiring hundreds or even thousands of calculations. In such cases an alternative approach can be used to obtain a solution with a minimum of calculation.

Assume that it is desired to determine the maximum junction temperature of a rectifier under steady-state conditions with repetitive power pulses as shown in Fig. 7(A). The pulses have constant amplitude P_0, constant width t_P, and constant period τ. Consider a pulse n which occurs after steady-state conditions have been reached. Replace the pulses prior to n by a continuous level of power equal to the average pulse power P_{avg}.

$$P_{avg}=\frac{t_P P_0}{\tau} \tag{13}$$

This condition is shown in Fig. 7(B). The maximum junction temperature rise, T_{max}, at the end of the $n+1$ pulse can then be calculated by the methods given previously:

$$T_{max}=\left[\frac{t_P R_T}{\tau}+\left(1-\frac{t_P}{\tau}\right)\times r_{(\tau+t_P)}-r_{(\tau)}+r_{(t_P)}\right]P_0 \tag{14}$$

which may also be solved for P_0:

$$P_0=\frac{T_{max}}{\frac{t_P R_T}{\tau}+\left(1-\frac{t_P}{\tau}\right)r_{(\tau+t_P)}-r_{(\tau)}+r_{(t_P)}} \tag{15}$$

The junction temperature at the beginning of the nth pulse is higher for the conditions of Fig. 7(B) than for those of Fig. 7(A). Thus the preceding calculation gives a worst-case calculation for the junction temperature. The calculation can be carried out for successive pulses beyond the $n+1$ pulse giving lower and more accurate values for T_{max}. However, the equation given for the $n+1$ pulse is usually accurate within 3%, which is sufficient for most practical applications.

Example

Problem: Calculate the maximum pulse amplitude for conditions the same as given in the previous problem.

Solution: For this problem $\tau=200$ sec, $t_P=50$ sec, ambient temperature $=40$ C, $T_{max}=200-40=160$ C. From equation 15 and Fig. 3

$$P_0=\frac{160}{\frac{(50)(5.1)}{200}+\left(1-\frac{50}{200}\right)4.36-4.15+2.78}=50.5 \text{ watts}$$

This compares with 51.2 watts obtained by the iterative solution and indicates that the solution is conservative and reasonably accurate.

Overload Following Continuous Duty

Another variation of the superposition approach can be used for overload calculations on semiconductors. A common case is that in which an overload is applied immediately following a steady-state load at less than the maximum rating of a semiconductor. The variation of junction temperature with time is defined by

$$T_J-T_A=P_{CD}R_T+(P_{OL}-P_{CD})r_T \tag{16}$$

where

T_J = junction temperature, C
T_A = ambient temperature, C
P_{CD} = continuous duty heating in semiconductor, watts
P_{OL} = heating on semiconductor during overload period, watts
r_T is at a given time after application of overload, C/watt

Example

Problem: Assume that the semiconductor device for which thermal resistance characteristics are shown in Fig. 3 (4- by 4-inch fin) operates continuously with 10-watt losses in a 55 C ambient. What is the maximum load that can be suddenly applied for 1 minute without exceeding a junction temperature of 200 C?

Solution: Solving equation 16 for P_{OL} gives

$$P_{OL}=\frac{T_J-T_A-P_{CD}R_T}{r_T}+P_{CD}$$

and substituting the values given above, and extracted from Fig. 3, gives

$$P_{OL}=\frac{200-55-10\times5.1}{29}+10=42.4 \text{ watts}$$

With the use of the published specifications for this particular device, this value of power dissipation can be translated into terms of current.

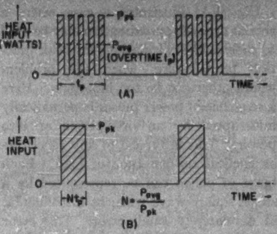


Fig. 9. Approximating a short train of power pulses with a single rectangular waveshape. A—Actual heat input. B—Equivalent heat input

Treatment of Irregularly Shaped Power Pulses

Approximation Method

The discussions thus far have assumed in every case that the input heat or power applied to the semiconductor junction was a step function, and that when power was plotted against a time base, a rectangular-shaped wave resulted. In most practical applications, the power pulse is not of this ideal shape for computation, and appropriate approximations must be made to convert the actual waveshape into a rectangular form if the subsequent calculations are to be made as outlined.

Fig. 8(A) illustrates the arbitrary waveshape of a power pulse that reoccurs at a period of τ sec and has a peak value of P_{pk} watts and a full-cycle average of P_{avg} watts. For the purpose of calculating peak junction temperature this waveshape can be approximated by the rectangular waveshape of Fig. 8(B), which is selected to have the identical values of peak power P_{pk} and average power P_{avg} as Fig. 8(A) by altering the pulse duration by a constant N to maintain the peak to average relationship. N is defined as the ratio of P_{avg} to P_{pk}.

This translation into rectangular pulses of power ensures a worst-case approximation since a rectangular pulse of power will always have an effect on temperature equal to or greater than the effect of any other pulse having the same peak and average power. In other words, a rectangular power pulse will raise the junction temperature higher than any other waveshape with the same peak and average values since it concentrates its heating effects into a shorter period of time, thus minimizing thermal losses during the pulse.

Fig. 9(A) illustrates a case where a similar type of approximation may be used to shorten the calculations for peak junction temperature, when the problem would otherwise be too laborious. Here a sequence of power pulses is periodically interrupted by an OFF period of zero power. This is typical of any repetitive or cyclical ON-OFF type of load. Each burst of pulses can be represented by a single square wave with the introduction of a relatively small error. This error will always yield a junction temperature higher than actual, and will thus provide a conservative application. In the equivalent waveshape shown in Fig. 9(B), the peak value of power P_{pk} is maintained as in Fig. 9(A). The duration of the equivalent rectangular waveshape is reduced to Nt_p where N is defined as P_{avg}/P_{pk}.

Example

The following is an example of conversion to rectangular wave power pulse.

Problem: Half-wave sinusoidal current flows through a high-current rectifier at 60 cycles per second. The full-cycle average value of this current is 50 amp (amperes). Approximate the heating effect of a sequence of four pulses of current by a single rectangular wave of power.

Solution:

$I_{pk}=\pi I_{avg}=50\pi=157$ amp
$V_{pk}=1.2$ volts at 157 amp from published specifications on the device
$P_{pk}=I_{pk}\times V_{pk}=157\times 1.2=190$ watts

Full-cycle average $P=50$ watts at 50 amp average current from published specifications on the device.

P_{avg} over 3.5 cycles (actual duration of heating) $=50$ watts$\times 4/3.5=57$ watts
$N=P_{avg}/P_{pk}=57/190=0.30$
$t=3.5$ cycles$\times 1/60$ cps$=0.0585$ sec
$Nt=0.3\times 0.0585=0.0175$ sec

Thus, the heating effect of the current can be approximated by a single rectangular pulse of power with an amplitude of 190 watts and a duration of 0.0175 sec.

Again, this is a conservative approximation. For the r_T characteristic in Fig. 3, the temperature rise due to the single pulse approximation is 85 C. With the more exact solution discussed in the section "Response to Equal-Amplitude Pulses," the temperature rise is calculated to be 82 C, indicating an error of less than 5%.

EXACT METHOD

Use of the superposition integral permits the exact calculation of the junction temperature as a function of time $T_J(t)$, provided that an analytical expression is available for both the r_T as a function of time, $r_T(t)$, and the applied power as a function of time, $P(t)$. The superposition integral may be used in either of two forms:

$$T_J(t)=P(0)r_T(t)+\int_0^t r_T(t-\tau)\frac{dP(\tau)}{d\tau}\,d\tau \tag{17}$$

$$T_J(t)=r_T(0)P(t)+\int_0^t P(t-\tau)\frac{dr_T(\tau)}{d\tau}\,d\tau \tag{18}$$

where $P(0)$ is the applied power at $t=0$, $r_T(0)$ is the r_T at $t=0$, and τ is the variable of integration.

It is found that an accurate analytical expression for the *C35* silicon controlled rectifier, shown in Fig. 3, is given by

$$r_T(t)=\frac{1+80t}{5+40t}$$

The chief disadvantage of the superposition integral is that only a limited number of waveforms can be expressed in terms of functions which can be integrated in equations 17 or 18.

Example

Problem: Assume that the power dissipation in a *C35* silicon controlled rectifier is given by

$P(t)=1{,}000t;\ 0\leq t\leq 0.1$ sec

$P(t)=0;\ t\leq 0,\ t>0.1$ sec

Determine the variation of junction temperature as a function of time and the value of the junction temperature at $t=0.1$ sec

Solution: Using equation 17 yields

$$T_J(t)=1{,}000\int_0^t\left(\frac{1+80t-80\tau}{5+40t-40\tau}\right)d\tau$$

Integrating yields

$$T_J(t)=1{,}000\left[2t-\frac{9}{40}\ln\,(1+8t)\right]$$

$T_J(0.1)=67$ C

This compares with 71.5 C obtained by approximating the actual power pulse with a rectangular power pulse.

Heat-Sink Design and Transient Thermal Resistance Curve

Since the heat sink is a major component in the heat transfer path between junction and ambient, its design substantially affects the r_T curve of Fig. 3. When the semiconductor manufacturer provides a device already mounted on a standard heat sink, he is in a position also to furnish r_T curves for the complete thermal system from junction to ambient. A typical set of this type of data will show a family of r_T curves for various velocities of cooling air.

However, when a semiconductor is manufactured and shipped to the user in a stud-mounted version, the manufacturer has no control over the ultimate heat sink and can provide data only on the heat transfer system between junction and stud, the part he manufactured. Fig. 3 shows these data as the curve identified as "cell mounted on infinite heat sink."

The equipment designer can use this curve in developing an r_T curve for the cell when mounted to a particular heat sink of his own design by means of a few simple calculations. The calculations assume the temperature throughout the heat sink as uniform even under transient loading, thus permitting the heat sink to be represented by a single time constant. This is a good assumption for fins of relatively thick cross section and fin effectiveness close to unity. This approach also assumes that the thermal capacitance of the heat sink is large compared to the thermal capacitance of the cell. By adding the calculated (or measured) r_T of the fin of the infinite heat sink curve for the cell, the applicable curve can be plotted for the complete thermal system. The method is best explained by an example illustrating how a specific heat sink design shown as the second curve in Fig. 3 (4- by 4-inch fin) can be calculated.

Example

Problem: A cell is mounted on a painted copper fin 1/16 inch thick and having 4 inches on a side. The fin is subjected to free convection air conditions. Find the r_T curve for a device mounted on this fin if the infinite heat sink curve in Fig. 3 represents the thermal characteristics of the stud-mounted semiconductor. Assume that the stud-to-fin contact resistance is negligible.

Solution: From fin design curves[8]

$h_r=0.005$ watt/inch^2 C

$h_c=0.005$ watt/inch^2 C

$h_{total}=h_r+h_c=0.010$ watt/inch^2 C

Fin thermal conductance $k=h\times A=0.01\times 4\times 4$ inches$^2\times 2$ sides$=0.32$ watt/C

Fin thermal resistance $R_F=1/k=1/0.32=3.1$ C/watt

Fin thermal capacity C
$=c\rho V=175$ watt-sec/lb C$\times 0.32$ lb/inch$^3\times 4$ inches$\times 4$ inches$\times 1/16$ inch
$=56$ watt-sec/C

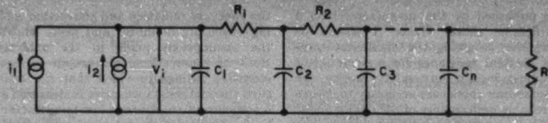


Fig. 10. Electrical analog for rectifier equivalent thermal circuit

Thermal R-C time constant $=3.1$ C/watt × (56 watt-sec)/C $=174$ sec

Equation of transient fin thermal resistance r_F

$$r_F=R_F(1-e^{-t/RC})=3.1(1-e^{-t/174})$$

The values of r_F are added to the infinite heat sink curve to secure the overall r_T of the system, as indicated in Fig. 3. Note that this fin has negligible resistance at periods of time one second or less after the application of power. In this area, it behaves as would an infinite heat sink which has no resistive element. The fin and the entire thermal system reach equilibrium at around 1,000 sec. Thereafter, the thermal capacity is no longer effective in holding down the junction temperature.

Appendix

The relationship between the heating curve and cooling curve can be derived for the electrical analog of the rectifier thermal equivalent circuit shown in Fig. 10. If the current generator i_1 produces a positive current step I, the input voltage v_i will be given as a function of time by $v_i=f(t)$. The limiting value of $f(t)$ for large values of t will be

$$\lim_{t\to\infty} v_i=\lim_{t\to\infty} f(t)=IR$$

(positive current step from i_1), where

$$R=\sum_{j=1}^{n} R_j$$

Similarly, if the current generator i_2 produces a negative current step $-I$, the input voltage v_i will be $v_i=-f(t)$, provided that the circuit elements are symmetrical. The limiting value of v_i for large values of t will be

$$\lim_{t\to\infty} v_i=-\lim_{t\to\infty} f(t)=-IR$$

(negative current step from i_2)

If the positive current step from i_1 is applied and the circuit is allowed to reach steady-state conditions, the superposition theorem indicates that a negative current step from i_2 will produce a variation in the input voltage given by

$$v_i=IR-f(t)$$

where the net input current (i_1+i_2) is zero and the limiting value of v_i for large values of t is zero. Use of the superposition theorem requires the circuit elements to be linear.

The preceding treatment applies to the thermal circuit as well as the electric circuit since they are analogs of each other. Current in the electric circuit is the analog of power in the thermal circuit; voltage in the electric circuit is the analog of temperature rise in the thermal circuit; and resistance and capacitance in electric circuit are analogs of thermal resistance and thermal capacitance in the thermal circuit.

Therefore the heating curve T_h and the cooling curve T_c of a rectifier are related by

$$T_c=PR_T-T_h$$

provided that the thermal equivalent circuit for the rectifier is linear.

The principal deviation from the ideal linear thermal characteristics would be caused by radiation effects which are proportional to the fourth power of the absolute temperature. However, tests on actual rectifiers have indicated a high degree of correspondence between the heating and cooling curves indicating that for most practical purposes radiation effects may be neglected.

References

1. RATING AND APPLICATION OF GERMANIUM AND SILICON RECTIFIERS, F. W. Gutzwiller. *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 75, 1956 (Jan. 1957 section), pp. 753–57.

2. THE RATING AND APPLICATION OF A SILICON POWER RECTIFIER, D. K. Bisson. "Conference on Rectifiers in Industry," *AIEE Special Publication T-93*, June 1957, pp. 135–42.

3. SILICON CONTROLLED RECTIFIER MANUAL. *Publication ECG-442*, General Electric Company, Liverpool, N. Y., first edition, 1960.

4. ELECTRICAL TRANSIENTS (book), L. A. Ware. G. R. Town. The Macmillan Company, New York, N. Y., 1954, p. 159.

◆

Discussion

E. J. Diebold and Werner Luft (International Rectifier Corporation, El Segundo, Calif.): Anyone who has struggled with this subject, to reconcile theory and experiment, must realize that there is no simple solution. The introduction of two novel and valuable principles by the authors of this paper is greatly appreciated. The concept of transient thermal resistance or impedance is not new: however, the expansion into a concept of a more general nature, to be manipulated similarly as a particular solution of a differential equation is very original, as is also the application of the superposition principle of mathematical solution to this more practical problem.

The method outlined seems to be of general validity. Mathematical proof should be given that the transient thermal resistance curve obtained from pulse load tests can be manipulated like a partial solution of a differential equation, and that the superposition principle applies for this solution.

To be acceptable as a true transient thermal resistance curve, the measured curve should exclude side effects which substantially change the shape of the curve. This separation may be difficult because the properties of materials change substantially as the heat pulses decrease in length and increase in magnitude. The ever present problem of measuring an instantaneous temperature rise becomes almost insurmountable for very short pulses. As long as the transient thermal resistance concept is used for simple 1-pulse investigations only, these inaccuracies are of no consequence. If, however, the curve is manipulated as a mathematical solution, the results obtained by this method may be entirely in the realm of numerical differences which are substantially less than the errors of tolerance of the measured magnitudes.

For each device and for each application of the method, the limits of validity of the curves and the probable error for each part of the curve must be stated. If the curves are drawn as bands (a method which is commonly used for electronic tubes), the temptation of reaching unwarranted conclusions is greatly reduced.

Interpretation of measured curves similar to the method described by this discusser[1] can give an indication of the trend of the curves. It may be necessary to implement the measured curves with such theoretical considerations to obtain a representative theoretical curve which can be manipulated as a mathematical solution and which yields results corresponding to subsequent experimental values.

REFERENCE

1. TEMPERATURE RISE OF SOLID JUNCTIONS UNDER PULSE LOAD, E. J. Diebold. *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 76, Nov. 1957, pp. 593–98.

F. W. Gutzwiller and T. P. Sylvan: In view of the discussion it should be re-emphasized that the entire approach discussed in this paper is based on actual measurements of junction temperature response to pulse loads. The thermal networks in Fig. 1 were included for the purpose of illustration and for explaining the basic assumptions on which the subsequent development is based. The validity of the presentation does not depend on a thermal analysis of the physical construction and materials in the semiconductor, or on the equivalent thermal network, beyond establishment that the thermal circuit is linear and bilateral, and that the heat source exists at only one plane inside the semiconductor. Under these assumptions, the use of superposition is valid.

As the discussers mention, difficulties do arise in actually measuring junction temperature response to very short load pulses due to instrumentation limitations. Colleagues of the authors have developed techniques that have acceptable accuracy for measuring the junction temperature within 200 microseconds after the end of the applied power pulse. With this measurement technique it is feasible to publish transient thermal resistance curves for pulse durations down to one millisecond, as shown in Fig. 3.

It should be emphasized that the major purpose of the analytical approach presented in this paper is to permit a convenient derivation of the practical ratings of power semiconductors applied under transient conditions. Hence, errors resulting from simplifying assumptions are not a serious limitation as long as they are on the conservative side. Some of the factors which must be considered in the evaluation of this method of analysis include:

1. In any case where a lumped, linear thermal equivalent circuit is used as the basis of analysis, the method of analysis given in the paper can also be applied if the conditions of validity are the same in both cases.

2. The transient thermal resistance curve should be considered as a rating curve when it is published by the manufacturer on a specification sheet. Hence it should be interpreted as giving the maximum transient thermal resistance for a given device and must include any safety factors the manufacturer thinks are necessary to ensure the reliable performance of the device under maximum heating conditions. It is not feasible to show a band for the thermal resistance curve, as this would require the manufacturer to test and reject for the minimum thermal resistance, which would serve no practical purpose.

3. Many of the deviations of the physical device from the ideal mathematical model result in a conservative rating. For example, as the peak current through a silicon-controlled rectifier or a diode rectifier cell is increased, a larger fraction of the power will be dissipated in the bulk material, contacts, and leads, since this power is proportional to the square of the current, whereas the power dissipated at the junction is approximately proportional to the first power of the current. This, however, results in a conservative rating provided that the transient thermal resistance curve is measured at a lower current level. Another example is the effects of radiation and convection cooling which are nonlinear and hence violate the assumptions made in the analysis. Both of these effects also result in conservative ratings provided that the transient thermal resistance curve is measured at a low-current (and hence low-temperature) level. The variation of thermal conducitvity of the various materials in the device at higher temperatures is negligible for rating purposes in the temperature range at which semiconductors are used.

4. One of the most serious limitations of any method of rating semiconductor devices under transient conditions is a consequence of the inherent electrical characteristics of the device. For example, localized imperfections and irregularities in the junction can cause localized heating which becomes more severe under high-current conditions and can result in a failure even though the transient rating of the gross device is not exceeded. A practical solution to this type of problem is for the manufacturer to run a portion of his quality control life tests under conditions corresponding to the minimum time given on the transient thermal resistance curve.

A Comparison of Digital Differential Analyzer and General Purpose Equipment in Guidance Systems

M. M. DICKINSON
NONMEMBER AIEE

QUALITATIVE comparisons are often made between a digital differential analyzer (DDA) and a general purpose (GP) computer for real-time control applications. Such comparisons consider static and dynamic accuracy, flexibility, equipment complexity, solution stability, influence of random errors, and ability to cope with discontinuities. They do not clearly point to a machine for solution of a problem since the DDA is found preferable in some characteristics while the GP excels in others. This ground is not to be covered.

This paper serves to record the physical characteristics of contemporary DDA and GP equipments actually designed for solution of a particular missile control problem. Qualitative comparison is made of the equipment programing difficulty and flexibility. For the problem considered the two equipments are seen to be competitive in most respects.

The Problem

One can think of the DDA as an outgrowth of an analog computer with two significant advances: first, the dependent variable is not limited to being time; second, the data are represented and processed discretely to permit greater precision. In the same sense the GP can be considered as an outgrowth of a desk calculator with advances resultant from its stored program and logical capability.

One might logically consider the application of a DDA in the missile guidance problem by formulating the differential equations describing the missile motion in an inverse square gravitational field and programing an analog of these equations on the machine. It should be expected that the DDA would yield an efficient solution to this problem although a moderately large machine (say, one of 100 integrators) would be required to permit consideration of all meaningful physical perturbations.

One might anticipate that the missile guidance problem had been dealt with by numerical analysts. Their work leads to a solution of the guidance problem through the use of polynomial equations in which variables are multiplied by constant coefficients and summed to derive guidance commands. Indeed the coefficients are derived from consideration of the physics of the situation, but the air-borne computing problem now manifests itself in multiplication and additive arithmetic processes. Machine solution can again be instrumented with the DDA, and although the process is less direct, a simpler instrumentation should be expected. Alternatively, it is now reasonable to consider instrumentation in machinery directly capable of multiplication and addition. The GP instrumentation which is suggested turns out to be competitive, as will be shown.

These comments suggest two generalizations: First, the control problem, although it may be analyzed through a mathematical analog of the physical situation, will commonly be manifested by polynomial equations with piecewise constant coefficients in the real-time control situation. Second, the effectiveness with which DDA and GP instru-

---

Paper 60-122, recommended by the AIEE Computing Devices Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., January 31–February 5, 1960, and re-presented for discussion only at the Fall General Meeting, Chicago, Ill., October 9–14, 1960. Manuscript submitted November 2, 1959; made available for printing July 15, 1960.

M. M. Dickinson is with the International Business Machines Corporation, Owego, N. Y.

Table I. Comparison of Physical Characteristics

	DDA	GP
Weight, lb	35	30
Volume, cubic ft	0.6	0.8
No. of components	3,840	3,989
Power, watts	60	78

mentation can be applied to a problem is critically dependent upon the mathematical details in which the problem is represented. Two computer users, each applying a control computer to a problem of a common class, could easily prefer a DDA on the one hand and a GP on the other, as a result of a slightly different approach in stating the problem. Where the user formulates the problem in terms of differential equations, the DDA is attractive; where he formulates in terms of polynomial expansions, the GP becomes more attractive.

The problem considered is the air-borne real-time control computation for a digital inertial missile guidance system. The calculations are piecewise continuous, and they involve the solution of polynomial equations. The computer input is vehicle acceleration expressed in an inertial orthogonal co-ordinate system. Vehicle acceleration is integrated and compared with required velocity to establish a velocity error. The velocity error is then used to generate missile steering signals and engine cutoff command as the computer output.

Computer Characteristics

The DDA selected is a high-frequency delay line machine and contains 30 processors which operate in a serial binary fashion at a rate of 1,500 iterations/sec (per second). The machine is constructed of 1-megacycle transistor circuits and memory is realized through the application of magnetostrictive delay lines.

The general-purpose computer is a serial machine using modular arithmetic element organization and floated-head magnetic drum storage. Separate multiply and add-subtract elements permit simultaneous instrumentation of 543 multiplication and 6,400 addition operations/sec. The machine is realized through the use of 200-kc transistor circuits and a magnetic drum memory. It has a 2,000-word serial binary storage.

Both machines have been found comparable in problem solution; both satisfy users' performance specifications. Preliminary machine organization, circuit design, and computer application studies have been effected in the consideration of the two machine types. More recently, similar machines have been realized in hardware.

The DDA iteration rate of 1,500 iterations/sec is not equalled by the GP solution rate of approximately 20 solutions/sec. The DDA rate is established by the maximum ratio of the time rate of change of problem variables to the corresponding variable resolution. For example, a 30,000-ft/sec (foot/sec) velocity range and 20-ft resolution requirement demand (30,000 ft/sec)/(20 ft/increment), or 1,500 increments/sec.

The GP rate is normally established by the bandwidth characteristics of the guidance system-missile dynamic control loop; that is, a control loop with a 5-cps (cycle/sec) bandwidth is found to behave satisfactorily when controlled by command signals generated at a rate of 20 times/sec.

For a control system with a 10-cps bandwidth, the GP would require twice the capacity, while the DDA would not necessarily be changed. Hence, those situations in which dynamics are critical have been popularly connected with DDA instrumentation.

Cost

The physical characteristics of the DDA and GP equipment are summarized in Table I, which includes cost factors most meaningful in equipment selection for air-borne application. Dollar cost should be proportional to component count. The relative value of the numbers tabulated points to no significant difference between the systems.

It should be recognized that the serial GP instrumentation which has been characterized in Table I is not conventional in its functional organization. It need not be for the class of control problems of which the missile guidance problem is characteristic. To realize an equipment competitive in cost with the DDA, and consistent with the problem requirements, the machine has no direct division capability. Programing is on a relative address basis requiring that the location of all data be carefully considered in the development of the program. As in the DDA, inputs are received on an incremental basis, accumulated external to the fixed program section of the machine, and the overflows of these accumulators are treated in the arithmetic element. It is under the constraints suggested by these characteristics that the following comments on relative programing difficulty and flexibility apply.

Programing of the DDA equipment involves scaling and designation of *dy* and *dx* inputs to each of the 30 processors contained in the machine. No significant restrictions are imposed on the interconnection of the processors in the machine organization selected. In the case of the GP equipment, programing involves the preparation of a 1,500-word 2-address program and the location of constants and intermediate data on the magnetic drum in such fashion as to permit utilization of the machine arithmetic element. The machine employs fixed-point arithmetic, and accordingly the programer must scale all data involved in the calculation. The ratio of programing difficulty based on limited observation of programing work would indicate that the effort involved in GP programing is four times greater than that involved in DDA programing. Similarly, programing modifications are more readily effected on the DDA equipment than on the GP, since changes in the GP program may require reoptimization of the location of significant portions of the program and intermediate storage.

Both digital differential analyzers and general-purpose machines offer the advantage of considerable flexibility over analog computers, since their programs and interconnections may be changed to satisfy the requirements of various problems. The machines may also be called upon to execute operations other than arithmetic, such as terminal equipment control and logical decisions based upon input parameters. Even in digital differential analyzers, decision and servo elements have been developed which extend their usefulness in this regard. DDA's are limited, however, in their capabilities to perform such functions as table look-up operations and certain logical choices and, indeed, some of these operations may be impossible.

It should be understood that the GP computer, with its broader logical and arithmetic capabilities, is more flexible than the DDA in executing secondary systems assignments. Where secondary assignments can be considered significant, then the GP solution is attractive. For example, in addition to the calculations involved in missile inertial guidance, the computer can be used for system ground checkout, platform calibration, and platform erection with nominal increase in machine memory. It is found that the same equipment can be time-shared between principal and secondary applications.

Selection

As stated, for the problem considered the two equipments are seen to be broadly competitive. Minor modifications in either the problem definition by the user or the computer technologies could easily reverse any decision made by consideration of any single cost factor or combination of cost factors suggested. In control applications where competitive physical designs can be identified, the GP has been considered preferable because of its greater flexibility. Programing difficulties are of minor importance unless a great number of modifications are anticipated.

Experience in the use of analog, DDA, and GP technologies in guidance systems has shown that there is no single optimum technology to be selected from the three basic alternatives. Some significant work has been done in the past 5 years in integrating these technologies in such fashion as to employ the best characteristics of each approach. Indeed, some of the results of this work are illustrated in the GP equipment discussed here. Further work must be effected in the integration of these technologies in order to advance the guidance system and control computer organization.

Hybrid computing equipments employing analog, DDA, and GP elements should be anticipated with increasing efficiency over the next 5 years. This trend will come to eliminate much of the question of choice between GP and DDA control system instrumentation, and result in a set of control computers for efficient application to the diverse control systems problems.

Error-Correcting Codes for Correcting Bursts of Errors

J. E. MEGGITT
NONMEMBER AIEE

IN DATA transmission systems that are subject to noise, it is found that errors do not occur randomly but in bursts. Consequently, much interest has centered lately on the problem of constructing suitable error-correcting codes.

In most of the codes described so far, difficulties have occurred with their implementation because different courses of action, depending on the nature of error burst that is detected, have to be followed by their decoders. This has made them expensive in equipment. The main purpose of the present paper is to show how some of these codes may be transformed to make their implementations simple.

Feedback Linear Sequential Systems

Feedback linear sequential systems whose feedback logic contains only modulo 2 adders may be conveniently described by modulo 2 matrix arithmetic. If such a system contains k elements, its state at any time may be described by a k-by-1 vector $\mathbf{x}$ whose elements are zero or one. The state at the next time is described by the vector:

$$\mathbf{Tx} \qquad (1)$$

where $\mathbf{T}$ is a k-by-k matrix all of whose elements are zero or one and the matrix multiplication is performed using modulo 2 arithmetic.

The behavior of such systems is determined, except in special circumstances, by the characteristic equation of degree k, that $\mathbf{T}$ satisfies.

$$F(\mathbf{T})=0 \qquad (2)$$

By a suitable choice of $F(\mathbf{T})$ it is possible to make the system go sequentially through 2^k-1 different states. This just means that the vectors $\mathbf{T}^r\mathbf{x}$ take all possible values for different r, except 0.

Such a characteristic equation will be described here by:

$$M(k\mathbf{T})=0 \qquad (3)$$

An excellent discussion of these ideas is provided in reference 1.

The same ideas, it transpires, are very useful in coding theory and the object of this paper is to explain their relevance.

Application to Codes

In the class of codes to be considered, data are transmitted in message blocks of n binary digits. In each block there are k check digits and n-k information digits. As each message is received, error correction is carried out.

If the digits in the message are a_1, a_1, $\ldots$ a_n, then the class of codes to be considered is defined by the equation:

$$a_1\mathbf{x}+a_2\mathbf{Tx}+a_3\mathbf{T}^2\mathbf{x}+\ldots+a_n\mathbf{T}^{n-1}\mathbf{x}=0 \qquad (4)$$

where $\mathbf{T}$ is a k-by-k matrix such as has been described, $\mathbf{x}$ is a k-by-1 vector and n is the lowest value of r for which

$$\mathbf{T}^r\mathbf{x}=\mathbf{x} \qquad (5)$$

These k equations 4 define k of the a's in terms of the other n-k. Thus only n-k digits are independent and these are taken to be the information digits in the message. The other k dependent digits are taken to be the check ones.

If the received message is a_1', a_2' $\ldots$ a_n', where the a's are the received and possibly incorrect digits, then it is possible to calculate:

$$a_1'\mathbf{x}+a_2'\mathbf{Tx}+\ldots+a_n'\mathbf{T}^{n-1}\mathbf{x}=\mathbf{z} \qquad (6)$$

This code is clearly capable of correcting any set of errors that produce different vectors $\mathbf{z}^2$, and the object of code design is to choose a $\mathbf{T}$ that ensures that the class of errors to be corrected does, in fact, produce different vectors $\mathbf{z}$.

When a matrix $\mathbf{T}$ has been chosen, equation 4 does not give the only set of coding equations that corrects the desired errors. It is possible to replace $\mathbf{T}$ in equation 4 by

$$\mathbf{U}=\mathbf{STS}^{-1} \qquad (7)$$

and $\mathbf{x}$ by $y=\mathbf{S}\mathbf{x}$, where $\mathbf{S}$ is a nonsingular matrix.

When this is done, the coding equations are transformed, but clearly the errors that such a code corrects are invariant under the transformation. The characteristic equation that $\mathbf{T}$ satisfies is also preserved under this transformation, and exactly as a linear sequential system is, in general, described by a characteristic equation,[1] so the code is determined in general by the characteristic equation that $\mathbf{T}$ satisfies.

Thus, when once a code is designed by the choice of a $\mathbf{T}$, it is possible to transform it into another form and convenient to transform it into such a form that is

---

Paper 60-654, recommended by the AIEE Computing Devices Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE North Eastern District Meeting, Providence, R. I., May 2–4, 1960, and re-presented for discussion only at the Fall General Meeting, Chicago, Ill., October 9–14, 1960. Manuscript submitted November 16, 1959; made available for printing August 10, 1960.

J. E. MEGGITT is with International Business Machines Corporation Research Laboratories, Winchester, England.

Vol.
Pt. 1
196

easily implemented. It is the object of this paper to describe how this is done.

Examples of Codes Determined by Characteristic Equations

EXAMPLE 1

If **T** is chosen to satisfy

$$M(k\mathbf{T})=0 \tag{8}$$

then the codes defined by equation 4 are such that single-error correction is possible, and by definition $n=2^k-1$. This is seen at once because a mistake in a_r leads from equation 6 to

$$\mathbf{z}=\mathbf{T}^{r-1}\mathbf{x} \tag{9}$$

By definition all the vectors $\mathbf{T}^{r-1}\mathbf{x}$ are different for different r. This set of codes was first noted by Abramson.[3]

As an example, it is possible to take

$$M(3\mathbf{T})\equiv\mathbf{T}^3+\mathbf{T}+1 \tag{10}$$

This gives a single-error correcting code of message length 7 with three check digits. It is then equivalent to one of the simple Hamming codes.

EXAMPLE 2

The above single-error correcting codes may be extended by adding a total parity-check equation. They are then capable of performing single and double adjacent error correction. For single errors, the code works as above. If, however, there is a double adjacent error in the rth and $(r+1)$th digits, this fact is indicated by the parity equation while the other equations lead to:

$$\begin{aligned}\mathbf{z}&=\mathbf{T}^{r-1}(1+\mathbf{T})\mathbf{x}\\&=\mathbf{T}^{r-1}\mathbf{x}^*\end{aligned} \tag{11}$$

Again, since by definition all the vectors $\mathbf{T}^{r-1}\mathbf{x}^*$ are different for different r, the value of r may be determined.

The extended code is still of the form of equation 4, but the new **T**, $\mathbf{T}^*$ say, is the $k+1$-by-$k+1$ matrix.

$$\mathbf{T}^*=\left[\begin{array}{c|c}\mathbf{T}&0\\\hline 0&1\end{array}\right] \tag{12}$$

where **T** is the k-by-k matrix used in the single-error correcting case.

Since **T** satisfies $M(k\mathbf{T})=0$, $\mathbf{T}^*$ satisfies

$$(\mathbf{T}^*+1)M(k\mathbf{T}^*)=0 \tag{13}$$

This class of codes was also noted by Abramson.

EXAMPLE 3

Melas[4] conceived the idea of extending the described simple additional parity check to a more elaborate set of equations. In this case, the factor $(\mathbf{T}^*+1)$ in equation 13 is replaced by $M_1(k_1\mathbf{T}^*)$ where $2^{k_1}-1$ is a factor of 2^k-1, and so the characteristic equation becomes

$$M_1(k_1\mathbf{T}^*)M(k\mathbf{T}^*)=0 \tag{14}$$

Such a code is capable of correcting various types of burst error. As in the examples, equation 2, the first factor in equation 14 indicates the nature of the burst, while the other factor positions it.

A simple example of such a code is the code described by

$$(\mathbf{T}^{*2}+\mathbf{T}^*+1)(\mathbf{T}^{*4}+\mathbf{T}^{*3}+1)=0 \tag{15}$$

This code has message length 15, six digits being check ones and it is able to correct single, double, and triple adjacent errors, together with two errors separated by one correct digit.

EXAMPLE 4

Another group of codes, and one that has been described in an unpublished paper by Fire, arises by replacing the factor $\mathbf{T}^*+1$ in example 2 by $\mathbf{T}^{*k_1}+1$ where k_1 is prime to 2^k-1.

One form of such a code is:

$$a_1\begin{bmatrix}1\\0\\0\\.\\.\\.\\0\\-\\\mathbf{x}\end{bmatrix}+a_2\begin{bmatrix}0\\1\\0\\.\\.\\.\\0\\-\\\mathbf{T}\mathbf{x}\end{bmatrix}+\ldots+a_n\begin{bmatrix}0\\.\\.\\0\\0\\.\\1\\\hline\mathbf{T}^{n-1}\mathbf{x}\end{bmatrix}=0 \tag{16}$$

Once more the first k_1 equations define the nature of the burst while the second k position it. Such codes are quite powerful. For example, the characteristic equation

$$(\mathbf{T}^{*5}+1)(\mathbf{T}^{*5}+\mathbf{T}^{*2}+1)=0 \tag{17}$$

gives a code of message length 155, with ten check digits. It is able to correct all burst patterns, 1, 11, 101, 111, 1011, 1111.

EXAMPLE 5

Bose and Chaudhuri[5] have given possible forms for general multiple-error correcting codes. From their theory it is possible to prove that a code with the characteristic equation

$$M(k\mathbf{T})M(k\mathbf{T}^3)=0 \tag{18}$$

corrects all double errors. Thus, for example, the code described by

$$(\mathbf{T}^4+\mathbf{T}+1)(\mathbf{T}^4+\mathbf{T}^3+\mathbf{T}^2+\mathbf{T}+1)=0 \tag{19}$$

has fifteen digits, eight of which are check ones, and is capable of correcting all double errors. Such a code may still be considered as a burst correcting code, although there are a large number of different burst patterns to be corrected.

Transformations

It is, therefore, seen that many useful codes are specified in terms of their matrix characteristic equations and, as has been stated, these equations completely determine their properties. Hence, sets of many apparently different codes, all with the same properties, can be constructed by using different matrices **T**, as long as all of them satisfy the same characteristic equation. Thus many transformations of the codes given by Abramson, Melas, and others exist, some of which are easier to implement than those which have previously been described.

There exists, in fact, a general procedure for implementing any code whose characteristic equation is given, and this will now be described.

The general characteristic equation has the form

$$\mathbf{T}^k+C_{k-1}\mathbf{T}^{k-1}+C_{k-2}\mathbf{T}^{k-2}+\ldots+C_1\mathbf{T}+C_0=0 \tag{20}$$

where the C's are zero or one. A convenient matrix that satisfies this equation is the associated matrix

$$\mathbf{T}=\begin{bmatrix}C_{k-1}&C_{k-2}&C_{k-3}\ldots C_1&C_0\\1&0&0\ \ldots 0&0\\0&1&0\ \ldots 0&0\\0&0&1\ \ldots\ .&.\\.&.&.\ \ldots\ .&.\\.&.&.\ \ldots\ .&.\\.&.&.\ \ \ldots 1&0\end{bmatrix} \tag{21}$$

The vector **x** of equation 4 can be chosen arbitrarily provided it is such that

$$F(\mathbf{T})\mathbf{x}\neq 0 \tag{22}$$

for any polynomial F of degree less than k. This may be ensured by taking

$$\mathbf{x}=\begin{bmatrix}1\\0\\0\\.\\.\\.\\0\end{bmatrix} \tag{23}$$

as may easily be seen.

This form of **T** is chosen because if the vector

$$\mathbf{T}^r\mathbf{x}=\begin{bmatrix}p_0\\p_1\\p_2\\.\\.\\.\\p_{k-1}\end{bmatrix} \tag{24}$$

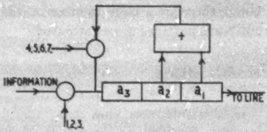


Fig. 1. Encoder

$$\mathbf{T}^{r+1}\mathbf{x}=\begin{bmatrix} C_{k-1}p_0+C_{n-2}p_1+\ldots+C_0p_{k-1} \\ p_0 \\ p_1 \\ \cdot \\ \cdot \\ \cdot \\ p_{k-2} \end{bmatrix} \tag{25}$$

Clearly, the operation of multiplying by **T** corresponds to a shift in a certain linear feedback shifting register.

Example 1

By using a matrix of the form 21, the Abramson code, equation 13 described by the equation

$$(\mathbf{T}+1)(\mathbf{T}^3+\mathbf{T}+1)=0 \tag{26}$$

which has seven message digits and which corrects single and double adjacent errors, has the form:

$$a_1\begin{bmatrix}1\\0\\0\\0\end{bmatrix}+a_2\begin{bmatrix}1\\1\\0\\0\end{bmatrix}+a_3\begin{bmatrix}0\\1\\1\\0\end{bmatrix}+a_4\begin{bmatrix}1\\0\\1\\1\end{bmatrix}+a_5\begin{bmatrix}0\\1\\0\\1\end{bmatrix}+a_6\begin{bmatrix}0\\0\\1\\0\end{bmatrix}+a_7\begin{bmatrix}0\\0\\0\\1\end{bmatrix}=0 \tag{27}$$

Explicitly, the coding equations are:

$$\begin{aligned} a_1+a_2+a_4&=0\\ a_2+a_3+a_5&=0\\ a_3+a_4+a_6&=0\\ a_4+a_5+a_7&=0 \end{aligned} \tag{28}$$

It is seen that each equation has the form of the previous one and this is important for the implementation to be described. This happens, of course, because the **T** used corresponds to a shift register. The same code as originally described by Abramson has the form

$$a_1\begin{bmatrix}1\\0\\0\\1\end{bmatrix}+a_2\begin{bmatrix}0\\1\\0\\1\end{bmatrix}+a_3\begin{bmatrix}1\\0\\1\\1\end{bmatrix}+a_4\begin{bmatrix}1\\1\\0\\1\end{bmatrix}+a_5\begin{bmatrix}1\\1\\1\\1\end{bmatrix}+a_6\begin{bmatrix}0\\1\\1\\1\end{bmatrix}+a_7\begin{bmatrix}0\\0\\1\\1\end{bmatrix}=0 \tag{29}$$

the last equation being the total parity check. The explicit coding equations here are obviously more complicated.

Example 2

The Melas code based on equation 15 produces the six coding equations:

$$\begin{aligned} &a_1+a_4+a_5+a_6+a_9+a_{10}=0\\ &a_2+a_5+a_6+a_7+a_{10}+a_{11}=0\\ &\cdot\\ &\cdot\\ &\cdot\\ &a_6+a_9+a_{10}+a_{11}+a_{14}+a_{15}=0 \end{aligned} \tag{30}$$

as may be verified.

Example 3

The Bose-Chaudhuri code based on equation 19 produces explicitly the eight coding equations:

$$\begin{aligned} &a_1+a_2+a_4+a_8=0\\ &a_2+a_3+a_5+a_9=0\\ &\cdot\\ &\cdot\\ &\cdot\\ &a_8+a_9+a_{11}+a_{15}=0 \end{aligned} \tag{31}$$

Implementation of Transformed Codes

For brevity, the encoding and decoding apparatus that is, in general, required will be described in detail only for the code 27. However, the method is absolutely general and the extensions will be shown.

When the message is transmitted, digits will be sent serially in order, starting with a_1. The last four digits (in general the last k) are chosen to be the check digits, and from equation 28 it is seen that they are defined by:

$$\begin{aligned} a_4&=a_1+a_2\\ a_5&=a_2+a_3\\ a_6&=a_3+a_4\\ a_7&=a_4+a_5 \end{aligned} \tag{32}$$

It is seen that each check digit is defined in terms of only the earlier digits, and this again is a consequence of using a **T** defined by equation 21.

Encoder

An encoder that imposed the conditions 27 is shown in Fig. 1. The squares indicate a shift register. Information flows into the register at times 1, 2, and 3 as shown by the timing pulses, while data are fed back at times 4, 5, 6, and 7 when the check digits are formed. The configuration is shown at time 4.

Extensions for codes 30, 31, and others are obvious. For 30, for example, the register is of length 9, and there are feedback connections from positions 1, 4, 5, 6, and 9.

Decoder

A decoder for the code 32 is shown in Fig. 2. The operation of the decoder is to store the received data in a shifting register of length 7, and to calculate the vector **z** of equation 6. The elements of **z** are defined by:

$$\begin{aligned} a_1+a_2+a_4&=z_1\\ a_2+a_3+a_5&=z_2\\ a_3+a_4+a_6&=z_3\\ a_4+a_5+a_7&=z_4 \end{aligned} \tag{33}$$

The z's are calculated in the same way that the check digits themselves are calculated, and they are calculated as the message is received and fed to a second shifting register. The decoder is shown at time 8, which is the same as time 1, times being measured from the arrival of the first digit in the message.

Henceforth, until time 4, the lower shifting register is fed back so that its contents are continually changing, while simultaneously, information leaves the main register. As this happens, the detection circuit seeks coincidences with one of the two patterns 0001, 0010. When a coincidence is found, the digit currently being output from the end of the

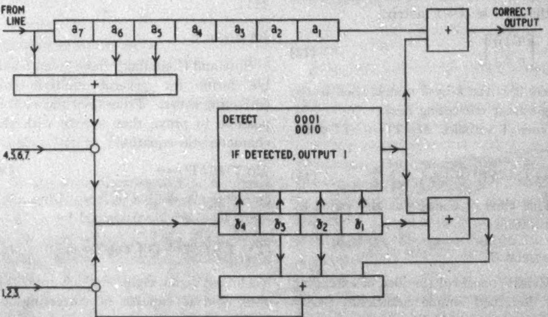


Fig. 2. Decoder

main register is inverted and corrected. Simultaneously, the lower register is altered to indicate that the burst pattern which is now left to be corrected is a simpler one, since the first digit of it has already been corrected.

This arrangement corrects, in this case, all single and double adjacent errors in the message.

Theory of Decoder

The role of the fed-back shifting register is to operate repeatedly on its contents **p** with $\mathbf{T}^{-1}$. Since the operation of **T** corresponds to the operation of a shift register, it is clear that the operation of $\mathbf{T}^{-1}$ can also be implemented by a shift register whose feedback connections are determined by the *C*'s of equation 20 and these are the connections shown.

The detection circuit is arranged to detect

$$\mathbf{p}=\mathbf{x}=(0001) \text{ and } \mathbf{p}=(1+\mathbf{T})\mathbf{x}=(0010) \qquad (34)$$

Thus, if there is a single error in the *r*th digit so that

$$\mathbf{z}=\mathbf{T}^{r-1} \qquad (35)$$

the detection circuit operates after $(r-1)$ shifts. However, at this time the *r*th digit is currently being output from the main register so that it is inverted as required.

The detection circuit also adds **x** to the lower register. This is accomplished just by adding an extra one to the feedback adder of the register, since **x** has the very simple form $\mathbf{x}=(0001)$.

Thus in the case of a single error $\mathbf{p}=\mathbf{x}+\mathbf{x}=0$ when the mistake is corrected, and the register becomes zero so that no further corrections take place.

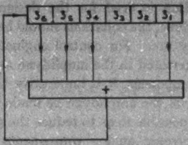


Fig. 3. Connections for Melas code, equation 15

For a double adjacent error in the *r*th and $(r+1)$th digits:

$$\mathbf{z}=\mathbf{T}^{r-1}(1+\mathbf{T})\mathbf{x} \qquad (36)$$

and again the detection circuit operates after $r-1$ shifts and corrects the *r*th digit. When this happens, the lower register contains $\mathbf{p}=(1+\mathbf{T})\mathbf{x}+\mathbf{x}=\mathbf{T}\mathbf{x}$ because of the addition of **x**.

The error now looks like a single error in the $(r+1)$th digit, so the correction of this digit takes place at the next digit time as previously described.

For a general code, the decoder is exactly similar. However, it is necessary to detect a different pattern in the detector for each burst configuration it is required to correct. Thus, if the code corrects, amongst other things, a burst pattern $q_1, q_2 \ldots q_r$, then it is necessary to detect in the lower register.

$$(q_1+q_2\mathbf{T}+q_3\mathbf{T}^2+\ldots+q_r\mathbf{T}^{r-1})\mathbf{x} \qquad (37)$$

and to correct all simpler burst patterns.

Thus, in the code of Melas, equation 15, which corrects error patterns 1, 11, 101, 111, the detection circuit must detect the four patterns

$$\mathbf{x},\ (1+\mathbf{T})\mathbf{x},\ (1+\mathbf{T}^2)\mathbf{x},\ (1+\mathbf{T}+\mathbf{T}^2)\mathbf{x} \qquad (38)$$

For a general code too, the main register has length equal to the message length. The lower register has length equal to the number of check digits, and the feedback connections to it are such that its successive contents are multipled by $\mathbf{T}^{-1}$. Fig. 3 illustrates these connections for the Melas code, equation 15.

Conclusions

A simple implementation has been given for codes described by a characteristic equation, and it has been observed that codes often arise in this form. Hence, this is a very powerful approach to the coding problem.

Many codes that have previously been described have been too complicated for them to be taken seriously. These, on the other hand, are all extremely economical in hardware. The implications of this simplicity are clearly enormous.

References

1. THE THEORY OF AUTONOMOUS LINEAR SEQUENTIAL NETWORKS, B. Elspas. *Transactions*, Professional Group on Circuit Theory, Institute of Radio Engineers, New York, N. Y., vol. CT-6, no. 1, Mar. 1959, p. 45.
2. MULTIPLE ERROR CORRECTION BY MEANS OF PARITY CHECKS, G. E. Sacks. *Ibid.*, Professional Group on Information Theory, vol. IT-4, no. 4, Dec. 1958, p. 145.
3. A CLASS OF SYSTEMATIC CODES FOR NON-INDEPENDENT ERRORS, N. M. Abramson. *Ibid.*, vol. IT-5, no. 4, Dec. 1959, p. 150.
4. A NEW GROUP OF CODES FOR CORRECTION OF DEPENDENT ERRORS IN DATA TRANSMISSION, C. M. Melas. *IBM Journal*, International Business Machines Corporation, New York, N. Y., vol. 4, no. 1, Jan. 1960, p. 58.
5. A CLASS OF ERROR CORRECTING BINARY GROUP CODES, R. C. Bose, D. K. Ray-Chaudhuri. *Information and Control*, New York, N. Y., vol. 3, no. 1, Mar. 1960, p. 68.

Amplidyne Circuit Analysis

G. W. BILLS
ASSOCIATE MEMBER AIEE

Synopsis: A method is shown for obtaining an analog of a typical amplidyne which has four control windings. Both the need for electronic analog differentiators and the possibility of circuits containing even-numbered amplifiers which give rise to unstable oscillations have been eliminated by the use of inverse inductances.[1,2] This makes it possible to have each amplidyne parameter represented by a computer component, which makes for ease of analysis and synthesis. Nonlinearities may be included in the analog and the analysis can be as rigorous as desired, limited only by the number of analog computer elements available. The method presented is applicable to circuit or machine containing mutual reactances.

ALTHOUGH the amplidyne has been thoroughly and competently analyzed for the cases where only one control winding is employed, little has been published on the analysis of multicontrol winding amplidynes.[3,4] Such amplidynes are widely used; elevator systems and and positioning controls are two applications that appear frequently in control analysis. In transient analysis of systems containing multicontrol winding amplidynes, the usual approach is simply to add directly together all the control winding signals. This neglect of mutual effects between control windings gives erroneous time constants for the amplidyne and should be avoided.

Since the transient analysis of complex amplidyne circuits is extremely difficult

---

Paper 60-1006, recommended by the AIEE Computing Devices Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Pacific General Meeting, San Diego, Calif., August 8–12, 1960. Manuscript submitted November 4, 1959; made available for printing August 29, 1960.

G. W. BILLS is with North American Aviation, Inc., Downey, Calif.

The author wishes to express his thanks to Mr. Paul Lackey for suggesting the problem. This paper is a partial fulfillment of the requirements for the degree of Master of Science at the University of California, Los Angeles, Calif.

unless some form of digital or analog computer is used as an aid, it is the purpose of this paper to develop and present an adequate analog of a fairly complex amplidyne. Transient solutions may then be obtained by setting this analog on an electronic analog computer or digital differential analyzer.

Circuit Analysis

The first step in the amplidyne circuit analysis is to write equations expressing the transient electrical behavior of the amplidyne (Fig. 1) in a manner similar to that of Bower.[3] Voltages denoted by V are generated voltages whereas those denoted by E are impressed voltages. Self-inductances are expressed as L_n and mutual inductances as M_{mn}, where m refers to the circuit in which a voltage is induced by a current in circuit n. Voltages arising from moving conductors cutting a flux are similarly denoted by $N_{mn}I_n$.

All the equations are expressed in operational form, specifically Laplace transforms. Because amplidyne speed is assumed constant, the transient mechanical behavior of the amplidyne is not included in the equations.

$$(R_f+L_fS)I_f+M_{f1}SI_1+M_{f2}SI_2+M_{f3}SI_3+M_{fc}SI_c-M_{fd}SI_d-M_{fq}SI_q-M_{fs}SI_s=E_f \quad (1)$$

$$M_{1f}SI_f+(R_1+L_1S)I_1+M_{12}SI_2+M_{13}SI_3+M_{1c}SI_c-M_{1d}SI_d-M_{1q}SI_q-M_{1s}SI_s=E_1 \quad (2)$$

$$M_{2f}SI_f+M_{21}SI_1+(R_2+L_2S)I_2+M_{23}SI_3+M_{2c}SI_c-M_{2d}SI_d-M_{2q}SI_q-M_{2s}SI_s=E_2 \quad (3)$$

$$M_{3f}SI_f+M_{31}SI_1+M_{32}SI_2+(R_3+L_3S)I_3+M_{3c}SI_c-M_{3d}SI_d-M_{3q}SI_q-M_{3s}SI_s=E_3 \quad (4)$$

$$M_{cf}SI_f+M_{c1}SI_1+M_{c2}SI_2+M_{c3}SI_3+(R_c+L_cS)I_c-M_{cd}SI_d-M_{cq}SI_q-M_{cs}SI_s=E_c \quad (5)$$

$$-(N_{df}-M_{df}S)I_f-(N_{d1}-M_{d1}S)I_1-(N_{d2}-M_{d2}S)I_2-(N_{d3}-M_{d3}S)I_3-(N_{dc}+M_{dc}S)I_c-(R_d+N_{dd}+L_dS)I_d+(N_{dq}-M_{dq}S)I_q+(N_{ds}-M_{ds}S)I_s=V_d \quad (6)$$

$$(N_{qf}+M_{qf}S)I_f+(N_{q1}+M_{q1}S)I_1+(N_{q2}+M_{q2}S)I_2+(N_{q3}+M_{q3}S)I_3+(N_{qc}+M_{qc}S)I_c-(N_{qd}+M_{qd}S)I_d-(R_q+N_{qq}+L_qS)I_q-(N_{qs}+M_{qs}S)I_s=V_q \quad (7)$$

$$-M_{sf}SI_f-M_{s1}SI_1-M_{s2}SI_2-M_{s3}SI_3-M_{sc}SI_c+M_{sd}SI_d+M_{sq}SI_q+(R_s+L_sS)I_s=E_s \quad (8)$$

These equations describe the voltage drops in the circuits of the amplidyne and can include effects due to saturation and brush shifts away from the neutral position. Considerable simplification of these equations is possible and desirable. If information is need for the design of an amplidyne, the equations should be solved as they are. For control engineers who are interested in the amplidyne as a control component, it is imperative that the analog of the amplidyne be made as simple as possible so as to reduce the number of electronic analog components necessary to represent it.

Amplidynes are normally operated so that brushes are set close to the neutral zones and pairs of brushes are usually set 180 degrees apart. For these conditions, it can be shown that the mutual effect between the direct and quadrature circuits is zero so that $M_{dq}=M_{qd}=0$, and $M_{qf}=M_{q1}=M_{q2}=M_{q3}=M_{qc}=0$ because the control windings are not collinear with the quadrature coils.

If the direct-axis compensating winding is 100 per cent effective, $M_{df}=M_{d1}=M_{dc}=M_{d3}=M_{dc}=0$. Also, $N_{fd}=N_{1d}=N_{2d}=N_{3d}=N_{cd}=0$, and it should be remembered that M's are symmetrical, i.e., $M_{nm}=M_{mn}$. The internal construction of an amplidyne is such that mutual inductances between all the windings except those between the control windings and between the control and compensating windings are very small when brushes are set near neutral, so that $M_{sc}=M_{dc}=M_{qc}=M_{sf}=M_{s1}=M_{s2}=M_{s3}=M_{cf}$ $M_{c1}=M_{c2}=M_{c3}=0$. Therefore, equations 1 through 8 may be simplified to

$$(R_f+L_fS)I_f+M_{f1}SI_1+M_{f2}SI_2+M_{f3}SI_3=E_f \quad (9)$$

$$M_{1f}SI_f+(R_1+L_1S)I_1+M_{12}SI_2+M_{13}SI_3=E_1 \quad (10)$$

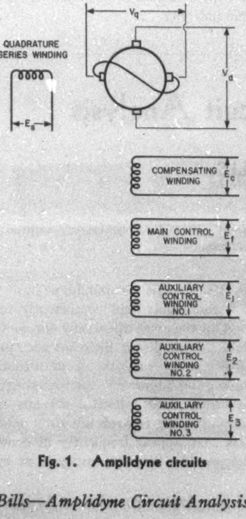

V_q
QUADRATURE SERIES WINDING
E_s
V_d
COMPENSATING WINDING
E_c
MAIN CONTROL WINDING
E_f
AUXILIARY CONTROL WINDING NO.1
E_1
AUXILIARY CONTROL WINDING NO.2
E_2
AUXILIARY CONTROL WINDING NO.3
E_3

Fig. 1. Amplidyne circuits

$$M_{2f}SI_f+M_{21}SI_1+(R_2+L_2S)I_2+M_{23}SI_3=E_2 \quad (11)$$

$$M_{3f}SI_f+M_{31}SI_1+M_{32}SI_2+(R_3+L_3S)I_3=E_3 \quad (12)$$

$$(R_c+L_cS)I_c=E_c \quad (13)$$

$$N_{dq}I_q-(R_d+N_{dd}+L_dS)I_d+N_{ds}I_s=V_d \quad (14)$$

$$N_{qf}I_f+N_{q1}I_1+N_{q2}I_2+N_{q3}I_3-N_{qd}I_d-N_{qs}I_s-(R_q+N_{qq}+L_qS)I_q=V_q \quad (15)$$

$$(R_s+L_sS)I_q=E_s \quad (16)$$

Further simplifications are possible when it is assumed that the quadrature series and quadrature armature circuits are connected in series and the compensating and direct armature windings are also connected in series. For these connections

$$V_q-E_s=V_q'$$

$$I_s=I_q$$

$$N_{qd}-N_{qc}=N_{qd}'$$

$$R_q+N_{qq}+N_{qs}=R_q'$$

$$L_s+L_q=L_q' \quad (17)$$

$$V_d-E_c=V_d'$$

$$I_c=I_d$$

$$N_{dq}+N_{ds}=N_{dq}'$$

$$R_c+R_d+N_{dd}=R_d'$$

$$L_d+L_c=L_d' \quad (18)$$

$$N_{dq}'I_q-(R_d'+L_d'S)I_d=V_d \quad (19)$$

$$N_{qf}I_f+N_{q1}I_1+N_{q2}I_2+N_{q3}I_3-N_{qd}'I_d-(R_q'+L_q'S)I_q=0 \quad (20)$$

It can be assumed that an amplidyne is connected as shown in Fig. 2 so as to control a drive motor in a positioning system. The black dots denote the polarity of each control winding, and loop currents have been indicated to aid in the analysis. Current I_7 is in the spillover circuit which limits the current in the control field and provides damping.

The entire control system is as shown in Fig. 3. This system was actually designed, built, and operated. Transient behavior of this system was determined before it was built by employing the amplidyne analog presented in this paper. Initial attempts to represent this system resulted in failure, and it was not until the analog, based on nodal analysis, was derived that a satisfactory solution was obtained. Except for the amplidyne, the control components were represented in a straightforward fashion.

The drive motor may be represented by

$$V_m=(R_m+L_mS)I_6+V_B \quad (21)$$

$$V_B=K_BS\theta_m \quad (22)$$

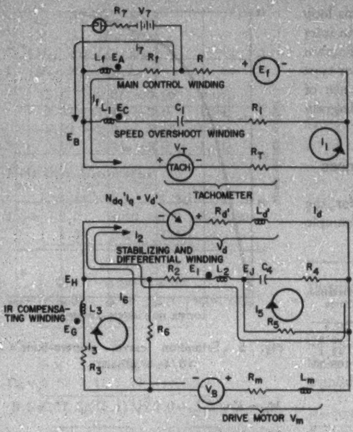

R7
V7
Lf
EA
I7
Rf
R
Ef
MAIN CONTROL WINDING
I1
L1
EC
C1
R1
EB
SPEED OVERSHOOT WINDING
VT
TACH
RT
I1
TACHOMETER
Ndq'Iq = Vd'
Rd'
Ld'
Id
I2
STABILIZING AND DIFFERENTIAL WINDING
Vd
EH
R2
E1
L2
EJ
C4
R4
EK
IR COMPENSATING WINDING
L3
I6
EG
I5
R5
R6
I3
R3
VB
Rm
Lm
DRIVE MOTOR Vm

Fig. 2. Amplidyne connections

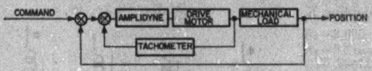

COMMAND
AMPLIDYNE
DRIVE MOTOR
MECHANICAL LOAD
POSITION
TACHOMETER

Fig. 3. Control system

$$T_m = K_4 I_6 = S(F_1 + J_1 S)\theta_m \tag{23}$$

$$\theta_m = \frac{K_4 I_6}{S(F_1 + J_1 S)} = \frac{K_4 I_6}{SF_1(1 + T_1 S)} \tag{24}$$

$$T_1 = \frac{J_1}{F_1} \tag{25}$$

$$V_T = K_T S \theta_m \tag{26}$$

An obvious approach in obtaining an analog representation is to apply block diagram algebra[5] to solve for the seven currents in Fig. 2. Once these currents are determined, the remaining unknown quantities may be easily determined by computer analysis.

Two difficulties can arise when mutual effects exist in electric circuits: 1. straightforward representation on an electronic analog computer would require the use of differentiating circuits which are not usually stable, and 2. avoiding the need for differentiation by equation manipulation leads to computer representations which include loops of even-numbered amplifiers which have no integrators. Such arrangements usually produce unstable oscillations.[6] To avoid such possibilities, the nodal form of equations can be used in block diagram algebra when mutual reactances are encountered. The only objection to this procedure is that the inverse self- and mutual inductances must be calculated, and this requires evaluation of determinants whose orders are equal to the number of inductances which are mutually coupled together. For this problem, the determinant order number of 4 does not present a difficult task.

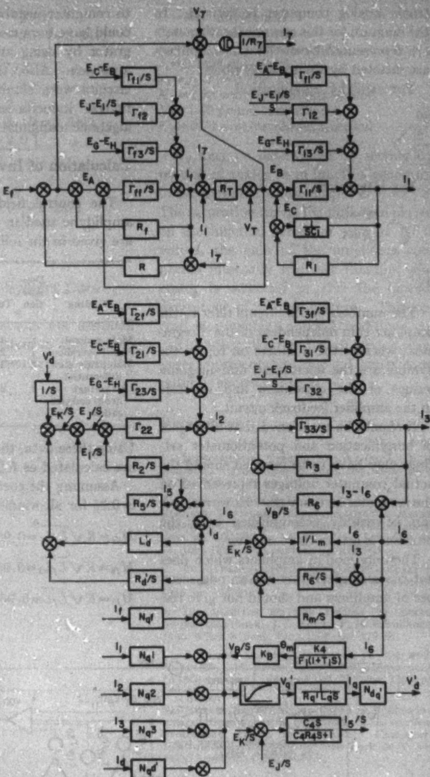

V7
1/R7
I7
EC-EB
Γf1/S
EJ-E1/S
Γf2
EG-EH
Γf3/S
I7
If
Ef
EA
Γff/S
RT
EA-EB
Γ11/S
EJ-E1/S
Γ12
EG-EH
Γ13/S
EB
Γ1f/S
I1
EC
1/SC1
Rf
I1
VT
R
I7
R1
EA-EB
Γ2f/S
EC-EB
Γ21/S
EG-EH
Γ23/S
Vd'
1/S
EK/S
EJ/S
Γ22
I2
E1/S
R2/S
I5
R5/S
I6
Id
Ld'
Rd'/S
EA-EB
Γ3f/S
EC-EB
Γ31/S
EJ-E1/S
Γ32
Γ33/S
I3
R3
I3-I6
VB/S
R6
EK/S
1/Lm
I6
I3
R6/S
Rm/S
If
Nqf
I1
Nq1
I2
Nq2
I3
Nq3
Id
Nqd'
VB/S
KB
θm
K4/F1(1+T1S)
I6
Vq'
Rq+LqS
Iq
Ndq
Vd'
EK/S
C4S/C4R4S+1
I5/S
EJ/S

Fig. 4. Amplidyne analog; signals are assumed to be added positively unless a negative sign appears in quadrant

Analog Derivation

By proper application of block diagram algebra and the use of inverse inductances, the analog of the control system shown in Fig. 3 can be derived. This analog is shown in Fig. 4.

Inverse inductances are denoted by Γ_{mn} and are found by inversion of the inductance matrix as follows:

$$\Gamma_{mn} = \frac{C_{mn}}{D}$$

$$D = \begin{vmatrix} L_{11} & L_{12} & L_{13} & L_{14} \\ L_{21} & L_{22} & L_{23} & L_{24} \\ L_{31} & L_{32} & L_{33} & L_{34} \\ L_{41} & L_{42} & L_{43} & L_{44} \end{vmatrix}$$

where

C_{mn} = cofactor of determinant D

Note that the saturation curve of the amplidyne is included in the analog diagrams.

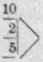
From the analog in Fig. 4, an electronic analog representation may be readily determined and scaled depending on the parameters of the amplidyne and the control system.

Specific Example

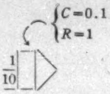
To illustrate the analysis procedure, the actual values for the components of a typical positioning control system, using the amplidyne shown in Fig. 2 will be given. Detailed calculations will be shown for the calculation of the inverse inductances.

By employing the amplidyne analog of Fig. 4, an actual setup for a Donner electronic analog computer is derived. In the diagram for this setup, potentiometers are represented by circles and amplifiers are denoted by the usual symbols.

For straight amplification:



For integrators:



The numbers on the inputs refer to the amplifier gain independent of the integration, whereas the numbers on top of the symbol are the microfarad and megohm values of the capacitors and resistors in the amplifier feedback circuit.

It should be mentioned that all values of amplification and potentiometer settings may have to be rescaled should the actual computer voltages represented on the computer be too high. However, this can be readily accomplished once the problem is set on the computer.

The only loop of amplifiers which does not contain integrators has an odd number of amplifiers and should not give rise to computer instability. Even this loop could have been made to include an integrator by using another block diagram approach. Also, the differentiator-type circuits were eliminated by the use of inverse inductances or block diagram algebraic manipulations.

Calculation of Inverse Inductances

The control field parameters for the amplidyne used in this specific example are given in the following table:

Winding	Designation	Turns	Inductance, Heurys	Resistance, Ohms
Main control	f	400	5.6	49.8
Speed overshoot	1	200	1.41	42.4
Stabilizing and differential	2	300	3.15	108.2
IR (voltage drop) compensation	3	200	1.41	43.8

Using these data, the inverse inductances are calculated as follows:

Assuming the coefficient of coupling K is 0.95 for all windings

$$M_{f1}=K\sqrt{L_fL_1}=0.95\sqrt{(5.6)(1.41)}=2.67$$

$$M_{f2}=K\sqrt{L_fL_2}=0.95\sqrt{(5.6)(3.15)}=3.99$$

$$M_{f3}=K\sqrt{L_fL}=0.95\sqrt{(5.6)(1.41)}=2.67$$

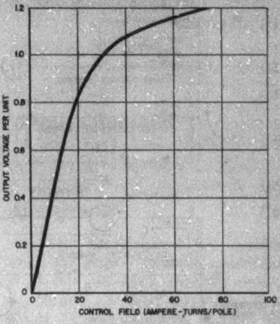
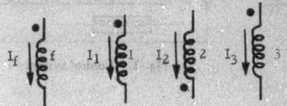


Fig. 5. Saturation curve, ampere-turns × 10/4π = gilberts

$$M_{12}=K\sqrt{L_1L_2}=0.95\sqrt{(1.41)(3.15)}=2.0$$

$$M_{13}=K\sqrt{L_1L_3}=0.95\sqrt{(1.41)(1.41)}=1.34$$

$$M_{23}=K\sqrt{L_2L_3}=0.95\sqrt{(3.15)(1.41)}=2.0$$


I_f
f
I_1
1
I_2
2
I_3
3

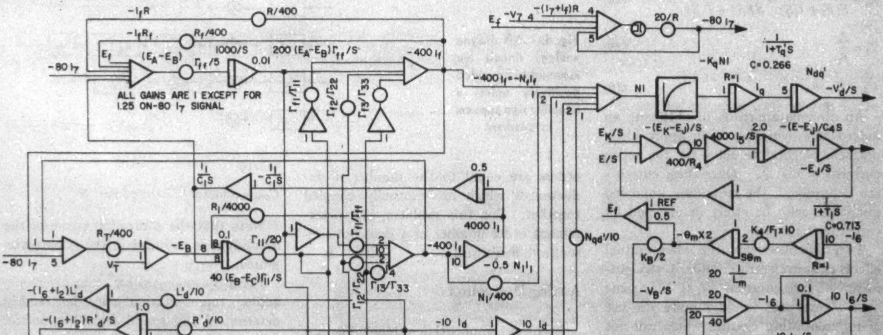

ALL GAINS ARE I EXCEPT FOR 1.25 ON -80 I7 SIGNAL
ALL GAINS ARE I EXCEPT LINE WITH Γ13/Γ11 POT WHICH IS 2.5
-80 I7
-IfR
R/400
-IfRf
Rf/400
(EA-EB)
Γff/5
1000/S
0.01
200 (EA-EB)Γff/S
-400 If
-400 If=-NfIf
Ef -V7 4 -(I7+If)R 4
20/R
-80 I7
-KqNI
NI
R=1
Iq
1/(1+T'qS)
C=0.266
Ndq'
-V'd/S
EK/S
-(EK-EJ)/S
400/R4
4000 I5/S
2.0
-(E-EJ)/C4S
E/S
-EJ/S
Ef
I REF
0.5
-θm X2
Km/F1 X 10
1/(1+T1S)
C=0.713
-I6
KB/2
5θm
R=1
20 1/Lm
-VB/S
-I6
0.1
10 I6/S
Rm/2
10 I6/S
-10 (I6-I3)R6/S
R6
-10 I3/S
10 I3
-10 I6
-10 (I3-I6)
10 I3
0.1
-0.5 N3I3
Nqd'/10
N2/400
N2I2
I1/C1S
-I1/C1S
0.5
R1/4000
4000 I1
RT/400
VT
-EB
0.1
Γ11/20
40 (EB-EC)Γ11/S
-400 I1
-0.5 N1I1
N1/400
Γ11/Γff
Γ12/Γ22
Γ13/Γ33
-(I6+I2)L'd
L'd/10
-(I6+I2)R'd/S
1.0
R'd/10
-10 Id
10 Id
V'd/S
EK/S
10EJ/S
Γ22/10
200 (E1-EJ)Γ22/S
0.025
-10 I6
-400 I2
I5R5/S
4000 I5/S
4000 I5/S
I2R5
E1
0.4
R2/1000
R5/4000
R3/100
Γ23/Γ32
Γ13/Γ11
-10 (I3-I6)
R6
-100 I3
R6-10
-10 (EG-EH)
0.05
Γ33/20
100 (EG-EH)
Γ33/S
Γ23/2Γff
Γ23/2Γ22
-100 I3

Fig. 6. Donner electronic analog computer representation of amplidyne positioning system

$$\Delta=\begin{vmatrix} L_f & M_{f1} & M_{f2} & M_{f3} \\ M_{1f} & L_1 & M_{12} & M_{13} \\ M_{2f} & M_{21} & L_2 & M_{23} \\ M_{3f} & M_{31} & M_{32} & L_3 \end{vmatrix}=$$

$$\begin{vmatrix} 5.6 & 2.67 & -3.99 & 2.67 \\ 2.67 & 1.41 & -2.0 & 1.34 \\ -3.99 & -2.0 & 3.15 & -2.0 \\ 2.67 & 1.34 & -2.0 & 1.41 \end{vmatrix}=0.017$$

Note that $M_{nm}=M_{mn}$

$$\Gamma_{ff}=\frac{\begin{vmatrix} 1.41 & -2.0 & 1.34 \\ -2.0 & 3.15 & -2.0 \\ 1.34 & -2.0 & 1.41 \end{vmatrix}}{\Delta}=2.72824$$

$$\Gamma_{11}=\frac{\begin{vmatrix} 5.6 & -3.99 & 2.67 \\ -3.99 & 3.15 & -2.0 \\ 2.67 & -2.0 & 1.41 \end{vmatrix}}{\Delta}=10.7235$$

$$\Gamma_{22}=\frac{\begin{vmatrix} 5.6 & 2.67 & 2.67 \\ 2.67 & 1.41 & 1.34 \\ 2.67 & 1.34 & 1.41 \end{vmatrix}}{\Delta}=4.7035$$

$$\Gamma_{33}=\frac{\begin{vmatrix} 5.6 & 2.67 & -3.99 \\ +2.67 & 1.41 & -2.0 \\ -3.99 & -2.0 & 3.15 \end{vmatrix}}{\Delta}=10.7235$$

$$\Gamma_{f1}=-\frac{\begin{vmatrix} 2.67 & -2.0 & 1.34 \\ -3.99 & 3.15 & -2.0 \\ 2.67 & -2.0 & 1.41 \end{vmatrix}}{\Delta}=-1.7706$$

$$\Gamma_{f2}=\frac{\begin{vmatrix} 2.67 & 1.41 & 1.34 \\ -3.99 & -2.0 & -2.0 \\ 2.67 & 1.34 & 1.41 \end{vmatrix}}{\Delta}=1.2059$$

$$\Gamma_{f3}=-\frac{\begin{vmatrix} 2.67 & 1.41 & -2.0 \\ -3.99 & -2.0 & 3.15 \\ 2.67 & 1.34 & -2.0 \end{vmatrix}}{\Delta}=-1.7706$$

$$\Gamma_{12}=-\frac{\begin{vmatrix} 5.6 & 2.67 & 2.67 \\ -3.99 & -2.0 & -2.0 \\ 2.67 & 1.34 & 1.41 \end{vmatrix}}{\Delta}=2.2471$$

$$\Gamma_{13}=\frac{\begin{vmatrix} 5.6 & 2.67 & -3.99 \\ -3.99 & -2.0 & 3.15 \\ 2.67 & 1.34 & -2.0 \end{vmatrix}}{\Delta}=3.6411$$

$$\Gamma_{23}=-\frac{\begin{vmatrix} 5.6 & 2.67 & -3.99 \\ 2.67 & 1.41 & -2.0 \\ 2.67 & 1.34 & -2.0 \end{vmatrix}}{\Delta}=2.2471$$

Positioning Control System Data

The following constants are assumed for the positioning system in addition to those which have been given for the amplidyne windings.

$R=100.0$ ohms
$R_f=49.8$ ohms
$R_1=3000$ ohms
$R_2=108.2$ ohms
$R_3=43.8$ ohms
$R_4=50.0$ ohms
$R_5=3{,}300.0$ ohms
$R_6=0.2$ ohm
$R_7=22.0$ ohms
$R_T=90.0$ ohms
$R_d'=3.88$ ohms
$R_m=1.126$ ohms
$R_q'=5.65$ ohms
$L_d'=0.4$ henry
$L_m=0.05$ henry
$L_q'=1.5$ henrys
$T_q'=0.266$
$N_{qf}=400$
$N_{q1}=200$
$N_{q2}=300$
$N_{q3}=200$
$N_{qd}'=8$
$N_{dq}'=5$
$C_1=500$ microfarads
$C_4=1{,}000$ microfarads
$F_1=0.20$ foot/radian/second
$J_1=0.1426$ slug-feet^2
$T_1=J_1/F_1=0.713$
$K_g=1.88$ volts/revolutions/second
$K_t=1.19$ pound-feet/ampere
$K_T=0.74$ volt/revolutions/second

K_q may be determined by setting up portion of analog from function generator to the amplifier whose output is $-Vd'/S$. The saturation curve, Fig. 5, is then placed on the function generator and, as NI is varied, the gain of the amplifiers connected internally with the function generator is adjusted to give the correct magnitude of output voltage. This operation simulates a no-load condition.

The block diagram for an analog computer representation, for a Donner computer, is shown in Fig. 6.

References

1. INTRODUCTORY CIRCUIT THEORY (book), E. A. Guillemin. John Wiley & Sons, Inc., New York, N. Y., 1953, p. 377.
2. NODAL REPRESENTATION OF LARGE COMPLEX-ELEMENT NETWORKS INCLUDING MUTUAL REACTANCES, J. C. Siegel, G. W. Bills. *AIEE Transactions*, pt. III (*Power Apparatus and Systems*), vol. 77, 1958 (Feb. 1959 section), pp. 1226–29.
3. FUNDAMENTALS OF THE AMPLIDYNE GENERATOR, J. L. Bower. *Ibid.* (*Electrical Engineering*), vol. 64, 1945, pp. 873–81.
4. INTRODUCTION TO THE DESIGN OF SERVOMECHANISMS (book), J. L. Bower, P. M. Schultheiss. John Wiley & Sons, Inc., 1958, pp. 479–81.
5. A BLOCK DIAGRAM APPROACH TO NETWORK ANALYSIS, J. M. Stout. *AIEE Transactions*, pt. II (*Applications and Industry*), vol. 71, Nov. 1952, pp. 255–60.
6. HIDDEN REGENERATIVE LOOPS IN ELECTRONIC ANALOG COMPUTERS, L. G. Walters. *Transactions*, Professional Group on Electronic Computers, Institute of Radio Engineers, New York, N. Y., vol. PGEC, sect. 2:1, June 1953.

Logical Organization of the Honeywell H-290

JOSEPH J. EACHUS
NONMEMBER AIEE

THE Minneapolis-Honeywell Regulator Company's *H-290* is a general-purpose digital computer which is provided with input and output paths which make it particularly easy to relate the computer to sensing and control devices.

The computer is fully parallel, operating on an 18-bit word. Typical execution rates are additions, 700 per second, and multiplications, 1,200 per second. The machine is provided with a core memory of 4,096 words, plus 8,192 words of drum storage.

Within one meaning of the term, the *H-290* is "micro-programed." That is, there is a clearly defined control section, and a clearly defined controlled section, so related that the wiring of and the hardware content of the controlled section is unaffected by changes in the computer's command vocabulary. This relationship may be said to function with the controlled section being capable of performing, on receipt of control signals, any of a collection of microsteps, and it being the function of the controlling section to cause these microsteps to be taken in sequence so that the eventual result is the execution of a command in the conventional sense.

It was the design objective, in the *290*, to produce a controlled section which would permit the execution of any general-purpose computer command, permit ready communications outward to

---

Paper 60-971, recommended by the AIEE Computing Devices Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE-AIChE-IRE-ISA-ASME Joint Automatic Control Conference, Cambridge, Mass., September 6–9, 1960. Manuscript submitted April 25, 1960; made available for printing July 29, 1960.

JOSEPH J. EACHUS is with the Minneapolis-Honeywell Regulator Company, Newton Highlands, Mass.

other electric devices, permit ready communications inward from electric devices, provide communications to and from paper tape, to provide direct visual communications to the operator, and to permit the operator to communicate directly to the machine. It was required that no feature of this section place significant restraints on the nature of equipment with which it might communicate, and that no lack of ability prevent the performance, under appropriate control, of any standard arithmetic or logical operation. This paper reports the logical nature of the controlled section designed to meet these objectives.

Notation and Terminology

Capital letters are used in the following several senses:

1. When one of the letters *A* through *V* is followed by three digits, one inverting amplifier is designated. The same symbol is used as the logical (Boolean) variable that is realized by the amplifier. Amplifiers are associated in pairs, the first of the three digits, always 0 or 1, distinguishing between the two of a pair, and the last two digits designating the pair. If the letter is among the 1's *A* through *K*, each pair of amplifiers is cross-coupled to form a bistable circuit, one or the other amplifier being saturated and its mate cut off. In this case, as logical variables one or the other of a pair is a 1 and its mate a 0. The statement: "Amplifier *A*013 is cut off" is to be taken as interchangeable with the statement: "The logical variable *A*013 is a 1." Similarly, "*A*013 is saturated" and "*A*013=0" are equivalent.

2. When one of the letters *A* through *V* is followed by two digits, the pair of amplifiers is designated. The expression: "Transfer from *D*13 to *A*13" means, in full, if *D*013 is a 1, cause *A*114 to become a 0, but if *D*113 is a 1, cause *A*014 to become a 0. Not all pairs of amplifiers are cross-coupled, and in particular, those designated by the letter *M* are not. The expression: "Transfer from *M*07 to *B*07" means, in full, if *M*007 is a 1, cause *B*107 to become a 0, or if *M*107 is a 1, cause *B*007 to become a 0, but if both *M*007 and *M*107 are 0, do nothing which will alter the state of the *B*07 pair. Restraints will be observed later which prevent both *M*'s of a pair from being simultaneously 1.

When referring to a cross-coupled pair, a statement such as "*C*12 contains a 0" is to be taken to mean that *C*012=1 and *C*112=0, and *C*12 contains a 1 and that *C*012=0 and *C*112=1. The term "Set *E*01 to 1" means, analogously, to cause *E*001 to become 0.

3. A letter without numerical suffix is used to designate the set of amplifiers which are individually designated by that letter with suffixes. If these amplifiers are cross-coupled pairs, the term "register" will be applied to the collection.

4. The letters *Y* and *Z* are used to designate signals emanating from the control section which cause action in the controlled section. *Y* followed by two digits designates such a signal which is normally off and comes on to cause action. *Z* followed by one digit designates such a signal which is normally on, and goes off to cause action. As logical variable, a *Y* is normally 0, but sometimes becomes 1, and *Z* is normally 1, but sometimes becomes 0. It is the function of the control section to emit these signals in orderly sequence.

While the nature of the control section is not a subject of the present discussion, reference must be made to a set of 12 cross-coupled amplifier-pairs therein, *W*00 to *W*11, inasmuch as some transfers take place, both ways, between them and elements of the controlled section. It is, in fact, the state of the *W*-register that determines which *Y*'s and *Z*'s will be emitted by the control section.

5. At times when an accurate and lucid description of an element and its function can be conveyed in words, letter symbols for the element will be omitted. For instance, the "drum busy flip-flop" is so treated.

Autonomous Parts

There are three integral portions of the computer, each of which is capable of a certain amount of independent action, such action being carried out on receipt of a signal asking for it. They are:

1. The core memory, which will act on receipt of a "read" or "write" request signal. On the advent of a read signal the memory address corresponding to the contents of *E*00–*E*11 will be read, the result appearing in the *F* register. On the advent of a write signal, the contents of *F*00–*F*17 will be written at the address specified by the contents of *E*00–*E*11.

2. The drum memory, which acts in a manner similar to that of the core memory, but with *Q*00–*Q*12 supplying addresses and *V*00–*V*17 serving as data register. Due to the indeterminate length of time required for the drum to reach the angular position corresponding to a given address, a means of indicating whether or not the last-called-for action has been completed is supplied, in the form of the drum busy flip-flop. This flip-flop is set to 1 by either the read request signal or the write request signal. It is set to 0 by the actual execution of a read or write, and hence contains a zero when, and only when, the drum memory is capable of receiving a new request for action.

3. The paper-tape punch section. This section contains eight data flip-flops *U*00 to *U*07, the punch busy flip-flop *U*08, and the punch motor flip-flop. This section will take cognizance of any of three request signals, namely, turn on punch motor, turn off punch motor, and punch. The third of these will turn on the punch motor if not already on, and cause to be punched one character on paper tape, that character having holes and lack of holes in accordance with 1's and 0's contained in *U*00 to *U*07. If the tape has fewer than eight levels, the higher order *U*'s are ignored. This section is capable of handling either the Friden eight-level punch, which is clutched, or the Teletype high-speed punch, which is continuously rotating.

Registers

At this point it is appropriate to enumerate the registers, give them short names, and indicate the number of flip-flops in each. The reader is cautioned to avoid interpreting the names in a restrictive way since, while in most cases a register does perform the function implied by its name, in many cases it performs additional functions.

The registers are:

Letter Designation	Number of Flip-Flops	Name
A	20	accumulator
B	18	quotient
C	19	sequence
D	20	utility
E	18	memory address
F	18	memory local
G	18	display
H	18	output selection
I	18	input selection
J	18	output magnitude
K	18	input magnitude
Q	13	drum address
U	8	punch
V	18	drum local

Subsequently, there will appear a set of logical variables composed of the letter *T* followed by four digits. The first digit has value 1, 2, or 3, the second has value 0 or 1, and the last two range from 00 to 17. Two of these *T*'s, differing only on the second digit of suffix, are paired in the same way as are the logical variables representing a flip-flop. It will therefore cause no confusion to refer to registers

Vol.
Pt. 1
196

$T1$, $T2$, and $T3$, even though the electrical states representing these logical variables do not arise from flip-flops. In the case of $T1$ and $T2$ they arise from manually set switches or, alternatively from the points of relays, and in the case of $T3$, they are the outputs of a real time clock.

Logical Entities Other Than Registers

There are two logical networks which provide interrelations among the register, which are now described.

1. The transfer amplifiers comprise a set L of 36 inverting amplifiers, not cross-coupled, having many controlled paths into them, and a second set of 36 inverting amplifiers M, not cross-coupled, with many controlled paths out, there being a smaller set of controlled paths from the L's to the M's.

A typical subset of the L's realize the equations:

$$\bar{L}013=A013\,Y15+B013\,Y16+C013\,Y17+D013\,Y18+E013\,Y19+F013\,Y20+K013\,Y21+T1013\,Y44+T2013\,Y45+T3013\,Y46+V013\,Y60$$

$$\bar{L}113=A113\,Y15+B113\,Y16+C113\,Y17+D113\,Y18+E113\,Y19+F113\,Y20+K113\,Y21+T1113\,Y44+T2113\,Y45+T3113\,Y46+V113\,Y60$$

$$\bar{L}014=A014\,Y15+B014\,Y16+C014\,Y17+D014\,Y18, \text{ etc.}$$

For these L's having last two digits from 00 to 05, there is an extra term: $W---Y04$, where the dashes represent the same digits as the suffix of the L. It is a constraint placed on the control section that at most, one of the numerous Y's in these equations may be 1 at any time.

It is seen, then, that if none of the H's involved is 1, then every L is 1, and if some Y other than $Y60$ is 1, then one of each pair of L's is 0, and that if $Y60=1$, then one of each pair $L00$ to $L05$ is 0, and all L's from $L06$ to $L17$ are 1.

A typical pair of M's realize the equations:

$$\bar{M}013=(L013+Z2)(L113+Z4)(L012+Z5)(L014+Z6)$$

$$\bar{M}113=(L113+Z1)(L013+Z3)(L112+Z5)(L114+Z6)$$

$\bar{M}00$ and $M17$ have an "end around" input:

$$\bar{M}000=(L000+Z2)(L100+Z4)(L017+Z5)(L001+Z6)$$

$$\bar{M}100=(L100+Z1)(L000+Z3)(L117+Z5)(L101+Z6)$$

$$\bar{M}017=(L017+Z2)(L117+Z4)(L016+Z5)(L000+Z6)$$

$$\bar{M}117=(L117+Z1)(L017+Z3)(L116+Z5)(L100+Z6)$$

Table I. Functions Performed by Y's

Control Signal	Function Performed
$Y01$	transfer $M01$–$M16$ to $A01$–$A16$
$Y02$	transfer $D00$–$D18$ to $A01$–$A19$
$Y03$	transfer $M01$–$M17$ to $B01$–$B17$
$Y04$	transfer $W00$–$W05$ to $L00$–$L05$
$Y05$	set $C18$ to 0 and disable counting by C
$Y06$	transfer $M00$–$M17$ to $D00$–$D17$
$Y07$	transfer binary output of adder to $D00$–$D19$
$Y08$	transfer decimal output of adder to $D00$–$D19$
$Y09$	transfer $M00$–$M17$ to $E00$–$E17$
$Y10$	transfer $M12$–$M16$ to $F12$–$F16$
$Y11$	transfer $E01$–$E17$ to $F00$–$F16$ (right shift)
$Y12$	transfer $M00$–$M17$ to $G00$–$G17$
$Y13$	transfer $M00$–$M17$ to $H00$–$H17$
$Y14$	transfer $M00$–$M17$ to $J00$–$J17$
$Y15$	transfer $A00$–$A17$ to $L00$–$L17$
$Y16$	transfer $B00$–$B17$ to $L00$–$L17$
$Y17$	transfer $C00$–$C17$ to $L00$–$L17$
$Y18$	transfer $D00$–$D17$ to $L00$–$L17$
$Y19$	transfer $E00$–$E17$ to $L00$–$L17$
$Y20$	transfer $F00$–$F17$ to $L00$–$L17$
$Y21$	transfer $K00$–$K17$ to $L00$–$L17$
$Y22$	transfer $M00$ to $A00$
$Y23$	transfer $M00$ to $B00$
$Y24$	transfer $M00$–$M17$ to $C00$–$C17$
$Y25$	cause C to count one
$Y26$	transfer $M00$ to $F00$
$Y27$	turn on punch motor
$Y28$	turn off punch motor
$Y29$	transfer $D18$ to $A17$ and $D19$ to $A18$ and $A19$
$Y30$	transfer $D18$–$D19$ to $A18$–$A19$
$Y31$	transfer $A00$–$A19$ to $D00$–$D19$
$Y32$	set adder to add
$Y33$	set adder to subtract
$Y34$	set adder to decimal
$Y35$	set adder to binary
$Y36$	set $B01$ to 1
$Y37$	transfer $M17$ to $A18$–$A19$
$Y38$	transfer $M17$ to $W09$
$Y39$	transfer $M00$–$M07$ to $U00$–$U07$ and start punch motor
$Y40$	transfer $U08$ to $W09$
$Y41$	set $B00$ to 0
$Y42$	set $B00$ to 1
$Y43$	transfer $M12$–$M17$ to $W00$–$W05$
$Y44$	transfer manual register 1 to $L00$–$L17$
$Y45$	transfer manual register 2 to $L00$–$L17$
$Y46$	transfer time register to $L00$–$L17$
$Y47$	cause core memory to read
$Y48$	cause core memory to write
$Y49$	start paper-tape reader
$Y50$	set $K17$ to 1
$Y51$	transfer $N19$ to $W09$
$Y52$	transfer $M17$ to $F17$
$Y53$	transfer $M00$–$M17$ to $I00$–$I17$
$Y54$	transfer $E00$ to $W09$
$Y55$	transfer $M01$–$M11$ to $F01$–$F11$
$Y56$	transfer $C18$ to $W09$
$Y57$	transfer $M17$ to $A17$
$Y58$	stop primary clock
$Y59$	transfer $M00$–$M17$ to $V00$–$V17$
$Y60$	transfer $V00$–$V17$ to $L00$–$L17$
$Y61$	transfer $M00$–$M11$ to $Q00$–$Q11$
$Y62$	read drum
$Y63$	write on drum
$Y64$	transfer drum busy flip-flop to $W09$
$Y65$	start input converter
$Y66$	transfer $M12$ to $Q12$

From these equations it is seen that if $Y15=1$, $Z1=0$, and $Z2=0$, then each M takes on the state of the A having the same suffix, but if $Y16=1$, $Z1=0$, and $Z2=0$, each M takes on the state of the B having the same suffix. Here is thus a mechanism which will produce a copy of the contents of any of dozen sources at a common location, the M amplifiers.

If, instead of activating $Z1$ and $Z2$ and some Y, $Z3$, $Z4$, and $Y15$ are activated, each M takes on the state opposite that of the A having the same suffix, and thus can also be produced, at the M's, the complement of the contents of any of the source registers.

Use of $Z5$ or $Z6$ permits the generation, at the M's, of the contents of any source register displaced one position left or one position right, respectively.

If $Y15=1$, $Z1=0$, and $Z3=0$, every M with first digit of suffix 1 becomes a 1, and each M with first digit of suffix 0 becomes 0. That is, at the M's, a word of 1's is produced. Similarly, any input to the L's and $Z2=0$, $Z4=0$ will generate a word of 0's.

The mechanism whereby data generated at the M's are put into registers is manifested in the logical equations for those register flip-flops. For instance, the complete logical equations for $G13$ are:

$$\bar{G}013=G113+M113\,Y12$$

$$\bar{G}113=G013+M01\,Y12$$

The first term on the right in each of these two equations is an indication of the stabilizing cross-coupling. If $Y12=0$, or if $M113=0$ and $M103=0$, both these equations are satisfied by either $G013=0$, $G113=1$, $G113=0$. If, however, $Y12$ becomes 1 simultaneously with $M113$ or $M013$, only one of the two conditions $G013=0$, $G113=1$ and $G013=1$, $G113=0$ satisfy both equations. It is seen, then, that if $Y12=1$, the operation "transfer $M13$ to $G13$" is performed.

$Y12$ appears in every G equation, and consequently its becoming a 1 causes the operation: transfer M to G.

Registers A, B, C, D, E, F, G, H, I, J, Q, U, and V have similar transfer means from M, but in not all cases does one Y cause all 18 bits to be transferred. For example, $Y01$ causes the transfer of $M01$–$M16$ to $A01$–$A16$, $Y22$ causes the transfer to $M00$ to $A00$, and $Y57$ causes the transfer of $M17$ to $A17$, and $Y37$ causes the transfer of $M17$ to both $A18$ and $B19$. All such separations may be seen in the itemization of functions performed by Y's in Table I.

Consider now the effect of $Y16=1$, $A1=0$, $Y01=1$, $Y22=1$, $Y57=0$, all other Y's and Z's inactive. No M with first suffix 0 may become 1, and hence no A flip-flop may be changed from containing a 1 to containing a 0. On the other hand, an M with first suffix 1 may become a 1 if the B with the same suffix is 1, that is, if the corresponding B flip-flop contains a 1. The net result is that A will finally contain a 1 at any position in which it originally had a 1, and in addition at any position at which it originally had a 0 if the corresponding position of B contains a 1. That is, there has been produced in A the logical sum of the original contents of A and the contents of B.

The effect of $Y16=1$, $Z2=0$, $Y01=1$, $Y22=1$, $Y57=0$, all other Y's and Z's inactive, is similar, except the result in A is 0 at any position where either A or B had a 0, that is, to produce in A the logical product of original contents of A and contents of B.

2. The adder network is a complex logical network which provides elementary arithmetic ability. It has as inputs all amplifiers of A register and of B register. It has, in addition, three other steady-state logical inputs: $S0$, $S1$, and $S2$. Inasmuch as the control section emits only transient signals, the S's are realized in flip-flops which are set by control Y's. $Y33$ causes $S0=0$ and $S1=1$; $Y34$ causes $S0=1$ and $S1=0$; $Y34$ causes $S2=1$; and $Y35$ causes $S2=0$.

The adder network has two sets of 19 pairs of outputs, N and R. There will be at the N's the sum of the number contained in A and the number contained in B if $S0=1$, $S1=0$, and $S2=0$, with all numbers considered as binary. If, however, $S0=0$, $S1=1$, and $S2=0$, there will be at the N's the number in A minus the number in B, again with all numbers considered as binary. If $S0=1$, $S1=0$, and $S2=1$, the result appears at the second set of outputs R, and is the sum of the number in A and the number in B, this time numbers being expressed in the standard binary coded decimal fashion. Finally, if $S0=0$, $S1=1$, and $S2=1$, the result is at the R's, and is the number in A minus the number in B, numbers being expressed in binary coded decimal fashion. Results get back into the register array via controlled paths from N and R to D. A typical D pair has equations:

$$\begin{aligned}
\bar{D}005 &= D105+N105Y07+R105Y08+---\\
\bar{D}105 &= D005+N005Y07+R005Y08+---
\end{aligned}$$

$Y07$ and $Y08$ are in corresponding positions in all D equations, and hence $Y07$ is called "transfer binary output of adder to D," and $Y08$ "transfer decimal output of adder to D."

As previously indicated, equations for the adder network are quite complex. However, those for any two-bit positions, four apart, are alike, and it will suffice to give the equations for one group of four, say 04 to 07. Two letters, P and Q, not previously mentioned as logical variables, appear in these equations. However, they are fully defined by the equations and require no further description. It is readily seen that they play the role of borrows and carries. These equations are:

$$\begin{aligned}
\bar{P}004 &= B104Q103+S1A004B104+\\
&\quad S1A004Q103+S0A104B104+\\
&\quad S0A104Q103\\
\bar{P}104 &= P004\\
\bar{P}005 &= P105\\
\bar{P}105 &= B005P004+S1A105B005+\\
&\quad S1A105P004+S0A005B005+\\
&\quad S0A005P004\\
\bar{P}006 &= B106P105+S1A006B106+\\
&\quad S1A006P105+S0A106B106+\\
&\quad S0A106P105\\
\bar{P}106 &= P006\\
\bar{P}007 &= P107\\
\bar{P}107 &= B007P006+S1A107B007+\\
&\quad S1A107P006+S0A007B007+\\
&\quad S0A007P006\\
\bar{N}004 &= A104B104Q103+A004B004Q103+\\
&\quad A004B104Q003+A104B004Q003\\
\bar{N}104 &= N004\\
\bar{N}005 &= A105B105P104+A005B005P104+\\
&\quad A005B105P004+A105B005P004\\
\bar{N}105 &= N005\\
\bar{N}006 &= A106B106P105+A006B006P105+\\
&\quad A006B106P005+A106B006P005\\
\bar{N}106 &= N006\\
\bar{N}007 &= A107B107P106+A007B007P106+\\
&\quad A007B107P006+A107B007P006\\
\bar{N}107 &= N007\\
\bar{Q}007 &= P107+S2N106N107+\\
&\qquad\qquad S2N105N107\\
\bar{Q}107 &= Q007\\
\bar{R}004 &= N104\\
\bar{R}104 &= N004\\
\bar{R}005 &= N105Q007+N005Q107\\
\bar{R}105 &= R005\\
\bar{R}006 &= Q007N106+P107N007N005+\\
&\quad P007N107N106N105+\\
&\quad P107N107N006N105\\
\bar{R}106 &= R006\\
\bar{R}007 &= Q007N107+Q107N007N006N105+\\
&\quad P107N106N105N107\\
\bar{R}107 &= R007
\end{aligned}$$

Miscellaneous Special Properties

Although the registers and networks already described are adequate for all required logic and arithmetic, an advantage in time to perform multistop operations can be gained by providing some transfer paths which do not utilize the transfer amplifiers, and by providing a counting means which does not use the adder net. The latter is provided by making register C a binary counter, with $Y25$ the input which causes it to count. Since the between-stage coupling which makes it into a counter tends to react undesirably when a transfer from M to C takes place, a means of disabling the counting property is provided and controlled by $Y05$.

The direct-transfer paths provided are indicated in the itemized list of control functions in Table I.

It is at times necessary for the controlling section to originate a data bit, that is, to set a register flip-flop, in performing such operations as divide or square root. $Y36$, $Y41$, and $Y42$ provide means for such actions.

It is at times necessary for the controlling section to take one of two actions, the choice being determined by the state of some register flip-flop or the state of some external device. The ability to do this is provided by a number of controlled paths to $W09$, a flip-flop of the control section. $Y38$, $Y40$, $Y51$, $Y54$, and $Y64$ control such paths.

Means have been indicated ($T1$ and $T2$) whereby the operator can communicate directly with the machine. Direct communication in the other direction is provided by an adjunct to the G register. For each four binary digits of the G register there is an illuminated display capable of showing any one decimal digit number, and capable of showing any of the two decimal digit numbers 10 through 15. They correspond to the 16 possible combinations of states of four flip-flops, and thus "display" the contents of the G register. G has no other function.

Means have been indicated for communicating outward in the form of paper tape. Means of going from paper tape to the machine is provided by the relay-point alternative of $T2$, in which case the paper-tape reader operates the coils of the relays. A control signal, $Y49$, actuates the clutch of the reader.

Registers H, I, and J provide additional outward communicating paths. Outputs of the amplifiers themselves may be brought out or, alternatively, the amplifiers are capable of operating the coils of relays, the points of which are available externally. Register K provides another inward communicating path. Wires are brought out which may be set to K flip-flops from an external source.

In the event that the external source is a device requiring a start signal, a control signal $Y65$ is provided for that purpose. $Y65$ also sets $K17$ to 1, it being the expectation that on completion of its action the external device will set $K17$ back to 0. The path K to L, L to M, $M17$ to $W09$, permits the control section to take cognizance of the state of completion of the cycle of the external device.

If the external device is an analog-to-digital converter with a number of possible inputs, the I register is used to operate the switching network which will connect an individual input to the converter.

If it is desired that the computer supply input to a digital-to-analog converter which has output to one of many destinations, the J register will supply the digital input, and the H register control the switching of output to destination.

Example of Sequencing

To illustrate the manner in which the array that has been outlined is used to execute conventional computer commands, given in detail below are the actions performed by control to execute the "tally" order, which requires that a specified address be read, the number read augmented by 1 and then stored back at the address whence it came. It was assumed that, initially, the command has been read, and is therefore in the F register, but no further action has taken place. C will contain the address from which the command has just been read, and the last stop of execution is to read the command at the succeeding address.

Step 1: $Y20$, $Z1$, $Z2$, $Y09$, $Y43$, $Y47$, $Y25$.

At this time, control has not yet been given the operation code of the command, and can therefore do only a beginning step common to all commands. $Y20$, $Z1$, $Z2$, will produce a copy of the command word at the M's. $Y09$ will put it in the E register, $Y43$ will put the operation code portion in $W00$–$W05$ of control, $Y47$ will cause the memory to read, and $Y25$ will cause the C register to augment its contents by 1.

Step 2: $Y17$, $Z1$, $Z2$, $Y06$.

This is C to L to M to D, creating a copy of the contents of C in D.

Step 3: $Y20$, $Z1$, $Z2$, $Y24$, $Y05$.

The memory read action called for in step 1 has put the operand in F. This step moves it to C.

Step 4: $Y25$.

This causes C to count.

Step 5: $Y17$, $Z1$, $Z2$, $Y10$, $Y26$, $Y52$, $Y55$, $Y48$.

This moves the new contents of C to F and causes the memory to write. E has not been changed since step 1, and therefore contains the correct address.

Step 6: $Y18$, $Z1$, $Z2$, $Y18$, $Y24$.

Although the action called for was completed at step 5, the content of C was data and not command address. This step restores in C the contents previously put aside in D.

Step 7: $Y17$, $Z1$, $Z2$, $Y09$, $Y47$.

This ending is common to all orders other than stops and jumps. It is C to L to M to E and read, which will bring the new command to F.

The tally command requires seven time-separated steps. The simplest commands take only three, while multiply requires 40 and square root 130.

Conclusions

The logical organization of the controlled section of a computer as described permits, with no wiring changes in the controlled section, wide variations of command vocabulary, including the possibility of single commands ordinarily considered prohibitively complicated. For example, one *H-290* order will scan a list of numbers, put the largest in the A register and store the location at which it was found at a special location in memory. However, a large vocabulary of complicated orders puts an immense requirement on control, and unless economical means of exercising control is available, such an organization is not to be recommended.

Transient Thermal Impedance of Semiconductor Devices

E. J. DIEBOLD
MEMBER AIEE

WERNER LUFT
ASSOCIATE MEMBER AIEE

Synopsis: In this paper the thermal behavior of the semiconductor device is referred to the junction. Devices consisting of multiple parts which are nonhomogeneous and with varying outlines are subdivided into homogeneous elements with simple outlines which are successively related to the junction. Thermal impedance calculations of cylindrical and spherical sectors are given numerically. Stationary thermal impedances provide the framework to assemble a valid succession of transient thermal impedances. The analysis is expanded for long heating cycles by a simple lumped-component approach with limits of validity.

SEMICONDUCTOR devices usually consist of a thin junction zone having a high, localized electrical and thermal resistance, within a comparatively large and thick body. During the operation of the device a large amount of heat is produced in this thin, plane zone. The heat flows by conduction into the remainder of the body which produces very little heat. Heat flowing away from the junction necessitates a temperature differential.

Heat flow from the junction into the surrounding mass of the body follows the laws of diffusion. Heating cycles which are shorter than the time required for heat diffusion through the full mass of the body, are treated with the concept of transient thermal impedance. Longer heating cycles may use the same condition but use a lumped-component approach. The ultimate limit is the stationary case (stationary thermal impedance).

In a previous paper,[1] the transient temperature rise of solid junctions was analyzed for different conditions. The present paper does not repeat this derivation but does take advantage of the formerly reached conclusions.

Pulses of increasing current, decreasing current, or constant current were correlated and equation 44 of reference 1 shows that such pulses behave similarly, hence a pulse of constant magnitude creates approximately double the temperature rise from a pulse with steadily increasing or steadily decreasing magnitude. Only pulses of constant magnitude are considered.

Successive approximations are used and a better method of correlating them has been developed. Every step is related to the junction.

It is assumed that all the heat of the device is generated only in the junction. This approximation of reality is very close.

The junction is of finite thickness but thin enough to permit a simple approximation of a plane homogeneous body in which the heat diffuses evenly, but only to one side of the plane and not through the plane. This assumption is also close to reality, except for point contact devices.

Heat escaping in both directions away from the plane is treated by calculating

---

Paper 60-68, recommended by the AIEE Semiconductor Metallic Rectifiers Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., January 31–February 5, 1960, and re-presented for discussion only at the AIEE Fall General Meeting, Chicago, Ill., October 9–14, 1960. Manuscript submitted November 2, 1959; made available for printing June 28, 1960.

E. J. Diebold and Werner Luft are with International Rectifier Corporation, El Segundo, Calif.

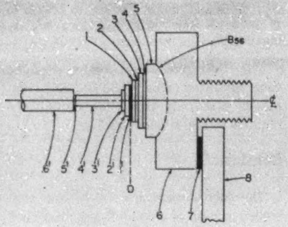


Fig. 1. Schematic representation of semiconductor device

the transient thermal impedance for each side and obtaining the resulting impedance from two parallel connected impedances. Heat is applied to the junction at a constant rate for a finite time. This approximation is apparently crude but was proved to be valid in the previous paper.[1]

Nomenclature

A = area, meters squared (m^2)
B = boundary
c = specific heat, watt-second per kilogram, degrees centigrade (ws/kg C)
C = thermal capacity, ws/C
d = thickness, m
D = diameter, m
γ = density, kg/m^3
k = thermal conductivity, w/mC
l = equivalent thickness (spherical section), m
O = origin of heat flow and reference point of temperature rise
p = propagation constant, m/$\sqrt{s}$
q = heat flow (at any point), w
Q = heat flow at origin (junction), w
r = radius of spherical sector, m
t = running time, seconds
T = time (fixed values), seconds
ΔT = time interval, seconds
θ = temperature rise, C
Θ = time constant, seconds
τ = temperature-rise constant, m^{2}C/w$\sqrt{s}$
u = normalized time
v = normalized transient thermal impedance
Z = stationary thermal impedance, C/w
ΔZ = incremental stationary thermal impedance, C/w
ζ = transient thermal impedance, C/w
ω = solid angle of spherical sector

Subscripts:

b = beginning
e = end
i = internal
j = junction
o = external
1, 2, 3 = specific numbered sections
w, x, y = numbered sections, general case
∞ = infinitely deep body (for diffusion)

Stationary Thermal Impedance

The stationary thermal impedance of a semiconductor device is the ratio of the temperature rise of the junction above the ambience, divided by the heat flow issuing from this junction. Heat flow may also be expressed as power dissipated in the junction.

The stationary thermal impedance can be calculated by subdividing the device into cylinders, cones, or other simple geometric bodies and determining the thermal impedance for each such section separately.

Transient Thermal Impedance

The transient thermal impedance of a semiconductor device is the ratio of the temperature rise of the junction after a given time, over the power dissipated in this junction, the only cause for this temperature rise, during the same time.

The transient thermal impedance is fictitious, but practically useful because it relates power dissipation and temperature rise in the junction to the heating-up time. This thermal impedance is meaningless for any other part of the body and for any other time. It depends to a certain extent on the time function of the heat generated in the junction.

The transient thermal impedance is always less than the steady-state thermal impedance. In the transient case, heat diffuses from the hot junction into the thermal mass of the cold parts, whereas in the steady state the heat must be dissipated to the space outside the device. Transient values determine the permissible pulse load of semiconductor devices under various conditions of initial temperature, pulse magnitude, and duration. Measurements of transient thermal behavior and failure limits can be analyzed for purely thermal or thermomechanical stress failure.

SECTIONAL SUBDIVISION OF SEMICONDUCTOR DEVICE

Heat generated in the junction of a semiconductor device diffuses through the solid in a diffusion wave which is first plane and then assumes a more spherical shape. Later the wave may be deflected into a plate or similar heat exchanger. The study of the transient thermal impedance for pulse behavior of semiconductor devices concerns mainly the initial stages but may consider the later stages by rather crude approximations. For the earlier stages only two approximations need to be considered: the plane approximation and the spherical diffusion which is found to be so flexible that it covers almost all practical cases. It is assumed that heat expands linearly into a conical body which is limited laterally by a thermally insulating cone, and in which the surfaces of equal temperature at equal times are geometrically similar parts of a sphere. The plane approximation is one particular case of the spherical approximation in which the opening angle of the cone is infinitely small and the distance from the center of the sphere to the flow volume is infinitely long.

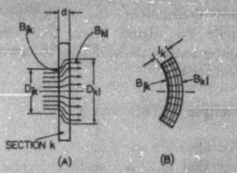


Fig. 2. Heat flow through plane disk (A) and equivalent spherical diffusion (B)

Subdivision into a plurality of sections is necessary because of the different materials which are used in semiconductor devices. Properties of the materials vary greatly and correct consideration of material constants is more important than precise coverage of dimensional variations. Thus the following analysis is based on the subdivision of a complex body into several sections which are made of different materials.

Fig. 1 shows a schematic representation of a semiconductor device without its housing. The body is assumed to have circular symmetry around the center line (CL). The different parts of the device between the junction and the cooling fin are indicated by numbers *1* through *8*. The number *0* is given to the junction. On the left-hand side, another heat-flow path is indicated by the numbers *1'* through *6'*. Practical analysis must cover both halves separately.

Subdividing the right-hand part assumes that the numbered sections are all made of different materials, except sections *5* and *6* which show abrupt change of shape. The boundary between section *5* and *6*, indicated by B_{56}, can be assumed to fit the spherical approximations for sections *5* and *6*. Section *7* is only the contact resistance between two solid bodies without mass; its transient and stationary thermal impedances are equal.

Fig. 2 shows the approximation of any of the sections *1* through *5* by a spherical body. Fig. 2(A) shows the section which is a circular, plane piece of material into which heat penetrates from the left-hand side in a fraction of its area and emanates on the right-hand side in a larger fraction of the area. In reality, comparing Fig. 1 to Fig. 2(A), the parts

1 through *5* have heat coming in from the left-hand side on a relatively small area but coming out on the right-hand side over the full area, but at unequal heat-flow density. This is studied by assuming that the diameter D_{jk} is given by the previous section *J* and the next diameter D_{kl} is not larger than D_{jk} plus two thicknesses *d* of the section, but the heat-flow density is uniformly high. In Fig. 2(B) the equivalent spherical diffusion with the boundary B_{jk}, the boundary B_{kl}, and the diffusion length l_k are shown. This approximation is easy to treat, has a stationary impedance equal to the original and a thermal mass which is approximately equal.

Fig. 3 shows another approximation of a section of a body which is similar to section *6* of Fig. 1. Fig. 3(A) shows the original with the boundary B_{56} at the left, which is assumed to be determined from section *5*. The boundary B_{67} is quite complex with heat transferred on the rim to a cooling fin and through the central stud to the hardware. A first approximation may be obtained, as in Fig. 3(B), in which the original boundary B_{56} assumes a ring shape, having the same area as in Fig. 3(A), and so does the outer boundary B_{67}. The spherical sector of Fig. 3(B) must have the same volume as in 3(A) and must present the same stationary thermal impedance as the original. Fig. 3(C) shows another approximation in which the thermal impedance is the same as in Fig. 3(B) because the volume, the areas of the boundaries, and the diffusion distances are equal.

Limits of Time and Stationary Thermal Impedance

If the device is subdivided into sections, each section must be correlated to the neighboring sections. This is done by means of fixed values of time and stationary thermal impedance. Stationary thermal impedances, similar to electrical impedances, may be added when connected in series and their reciprocals are added when connected in parallel. Transient thermal impedances, however, may not be added for sections connected in series, because the definition refers each thermal impedance to a junction, both by the heat produced and the temperature rise of the junction. By adding another section, later in the heat diffusion, the thermal impedance of the device, referred to the junction, can only be affected when the heat reaches the section, and then only in the measure that this section affects the junction. An approximation by which each section is treated as a new semiconductor device with the incoming section area a new junction, is not correct. On the other hand, if the thermal impedances and the elapsed time are simply added up, the results are also incorrect. These assumptions neglect the vital difference of the heat diffusion close to the source (steep wave) and the entirely different heat diffusion further away. Hence, an approximation is chosen in which the distance between the heat source and the section under consideration represents the real condition.

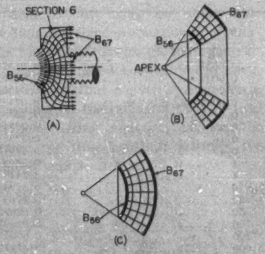


Fig. 3. Section of semiconductor device (A), first approximation of same section (B), final approximation of same section (C)

The thermal impedance of each new section is calculated by assuming that the section extends further than in reality, in both directions, towards the junction and away from the junction. The extension in the direction of the junction is made to obtain a stationary thermal impedance equal to the sum of the stationary thermal impedances of all the sections ahead of the section under investigation and between its initial boundary and a new fictitious junction. The transient thermal impedance is calculated from the new fictitious junction through the extension to the section and beyond. The heat diffusion wave reaches the end of the section when the transient thermal impedance reaches the total value of the stationary thermal impedance, referred to the end of the section. If the partial investigation is stopped at this point, the final thermal impedance is again equal to the sum of the stationary impedances from the real junction to the boundary under consideration. The diffusion time through the section is the difference of the time at which the transient thermal impedance is equal to the stationary thermal impedance at the end and the beginning of the section. The next section is treated similarly, by extending the section backwards to a new fictitious junction which creates an additional initial thermal impedance.

Briefly, each section is treated as a part of a homogeneous body which is heated by a properly located junction and into which the wave expands without obstacles. Transient impedance values are identical with partial stationary thermal impedances, their correlation establishes a time scale between impedance and heat pulse duration. When the heat wave has passed through the whole body, the final impedance is equal to the total stationary thermal impedance of the device.

The errors introduced with these assumptions are only tolerable if the investigation is applied to devices in which the incremental thermal impedance and mass show no excessive and sudden variations. The main error is that each analysis assumes that the thermal mass and conductivity per-unit volume of the preceding and following sections are identical with the section under consideration. For example, if a section with low specific heat follows a section of high specific heat, the transient thermal impedance rises much faster because the high specific heat of the preceding part is neglected.

Fig. 4 shows graphical representation of thermal impedances and times according to the foregoing assumptions. The origin O indicates the beginning of thermal diffusion when both time and transient thermal impedance are zero (the transient thermal impedance zero indicates that heat is diffusing immediately into a cold body). As heat diffuses at a steady rate into the body, the transient thermal impedance indicated by ζ_1 increases rapidly. The time is indicated by t_1. The stationary impedance of Z_1 of the first section is the limit for the transient thermal imped-

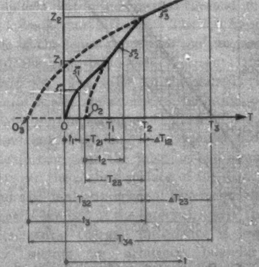


Fig. 4. Representation of nomenclature for thermal impedances and times

ance ζ_1; i.e., when the transient and the stationary thermal impedance of the first section are equal, it is necessary to assume that the heat diffusion has reached the point where the second section must be considered. The time at which the transient and stationary impedance of the first section are equal is T_1, the time limit for the first section.

Considering the second section, it is necessary to find a new origin O_2 which pertains only to the second section. This can be found if it is realized that the time T_{21} between the new origin and the time T_1, when the useful part of the approximation of the second section starts, is the time needed for the transient thermal impedance, referred to O_2, to reach the stationary thermal impedance Z_1. The running time t_2 for the second section, determines ζ_2, the transient thermal impedance. When it reaches the stationary impedance Z_2 (at the time T_{23} measured from O_2) the approximation for the second section reaches its limit of validity. The running time of the heat-diffusion wave in the second section is $T_2 = T_1 + \Delta T_{12} = T_1 + T_{23} - T_{21}$.

Fig. 4 shows also a third approximation in which the origin O_3 is put back by the time T_{32} required for the transient thermal impedance ζ_3 to reach the stationary thermal impedance Z_2 under the assumptions which were made for the equivalent body of the third section. The running time for the third section ΔT_{23} is again determined from the time required for the transient thermal impedance ζ_3 to increase from the stationary thermal impedance Z_2 to the new stationary impedance Z_3, and the time T_3, corresponding to the heat diffusion reaching the end of the third section, is computed by $T_3 = T_2 + \Delta T_{23} = T_2 + T_{34} - T_{32}$.

As plotted in Fig. 4, a solid-line plot of thermal impedance versus time is obtained. The method of successive approximations gives a correct interpretation of increasing impedance versus time, wherein each section which is further removed from the origin has a lesser increase of transient thermal impedance because of its remoteness, but an inherent rise of thermal impedance because of its physical characteristics (material and shape). Hence, ζ_2 rises steeply because of an inherently high thermal impedance but not as steeply as ζ_1, at the origin, which is due to closeness to the junction, although the inherent thermal impedance is lower.

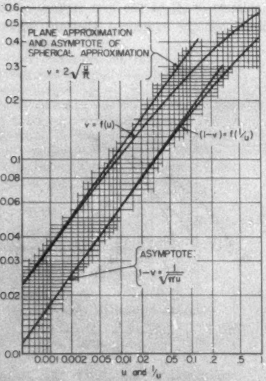


Fig. 5. Representation of equation $v = 1 - (1 - \Phi\sqrt{u})e^{u}$, shown as v versus u and $(1-v)$ versus $1/u$

Plane Approximation of a Homogeneous Section, Transient Thermal Impedance

Plane diffusion, as an approximation of this problem, is treated in reference 1. For short pulses the junction temperature is given in equation 20 of reference 1. Rewriting this equation in terms of transient thermal impedance gives:

$$\zeta = \frac{2\tau}{\sqrt{\pi} A}\sqrt{t} \tag{1}$$

This approximation is used in the immediate vicinity of the junction, as long as the heat-flow cross section remains essentially constant, as the heat wave progresses.

TIME SCALE

For each homogeneous layer, the transient thermal impedance, as calculated, is valid until the limit of equal transient and stationary thermal impedance is reached.

$$T = \left(\frac{\sqrt{\pi}}{2}\,\frac{AZ}{\tau}\right)^2 \tag{2}$$

where

$$Z = d/kA \tag{3}$$

or

$$T = \left(\frac{\sqrt{\pi}}{2}\,\frac{d}{p}\right)^2 \tag{4}$$

LIMITS OF TIME AND THERMAL IMPEDANCE

As shown for the spherical approximation, values of time and thermal impedance for successive sections can be matched by referring them back to a fictive junction.

Spherical Approximation of a Homogeneous Section

The device is subdivided into sections which are linked with each other by the limits of time and thermal impedance which are associated with the BEGIN and the END of the thermal-diffusion wave reaching and passing through each section.

Before the BEGIN the transient temperature rise of the particular section is negligible; after the END it is nearly equal to the stationary case.

Assuming that heat diffuses evenly into a homogeneous body which has the shape of a spherical sector, the lines of heat flow or heat diffusion will be straight and converge into a single point which is outside the body. All the areas of simultaneously equal temperature and heat-flow density are spherical and have this conversion point as a center. Lateral boundaries are formed by a generalized cone which is assumed to be a thermal insulator. The apex of the cone is formed by the conversion point of the flow lines. Fig. 3 shows a section of a practical body and its equivalent spherical sector, either as a ring sector or a solid sector. The numerical adaptation of any section of a real device to the spherical approximation is shown in Appendix I.

IMPEDANCE SCALE

The approximation by spherical diffusion is treated in reference 1, but there only a spherical sector with an opening angle of 180 degrees (hemisphere) is considered. The present analysis considers a general spherical sector of any solid opening angle. The basic equations presented in the former paper are written for the temperature rise of the junction (equation 51 in reference 1) and referred to heat-flow density and local tempera-

Table I

u	v
0.0001	0.01118
0.0009	0.03297
0.0016	0.04358
0.0049	0.07433
0.01	0.1035
0.09	0.2654
0.16	0.3292
0.49	0.4741
1	0.5724
4	0.7446
10	0.8294
20	0.8768
50	0.92099
100	0.94386
200	0.96021
5,000	0.97479
1,000	0.98217
2,000	0.98739
5,000	0.992022
10,000	0.994358

Vol.
Pt. 1
196

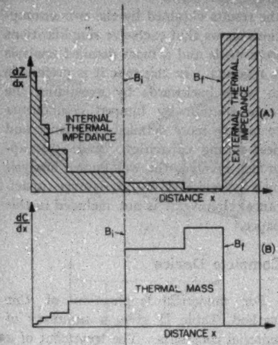


Fig. 6. Change in incremental thermal impedance (A) and capacity (B) versus distance from junction for semiconductor device

tures only and as such, are valid for any opening angle. Rewritten for the transient thermal impedance, the ratio of the temperature rises during the transient and steady-state condition for equal heat flow (of an infinitely large body) is equal to the ratio of thermal impedances for transient and steady-state heating. Equations 5 and 6 define the thermal impedances:

$$\zeta=\theta/q \tag{5}$$

$$\zeta_\infty=\theta_\infty/q=\frac{1}{\omega r_j k} \tag{6}$$

With the expressions for the solid angle and the junction radius defined by equations 21, 24, and 25, the transient thermal impedance of the infinite body is given by equation 7.

$$\zeta_\infty=\frac{1}{k\sqrt{\omega}\sqrt{A_j}}=\frac{Z_e\sqrt{A_e}-Z_b\sqrt{A_b}}{\sqrt{A_e}-\sqrt{A_b}} \tag{7}$$

TIME SCALE

Normalizing the time variable in the equation of thermal impedance can also be achieved by using the same time constant as defined in equation 52 of reference 1 and rewritten for the more generalized nomenclature defined in the foregoing. The time constant is given by equation 8, depending on the radius of the spherical junction. Using equation 25 for the radius, the time constant assumes the value of equation 9.

$$\Theta=(r_j/p)^2 \tag{8}$$

$$\Theta=\frac{\Delta Z^4}{\tau^2\left(\frac{1}{\sqrt{A_b}}-\frac{1}{\sqrt{A_e}}\right)^2\left(\frac{Z_e}{\sqrt{A_b}}-\frac{Z_b}{\sqrt{A_e}}\right)^2} \tag{9}$$

This time constant greatly simplifies the following equations. Physically it represents, for heat diffusing into an infinitely deep spherical sector, the time required by the transient thermal impedance to each 57.2% of the stationary thermal impedance of the same sector.

Numerical Values of Transient Thermal Impedance and Time

Time and transient thermal impedance are normalized, according to equations 10 and 11. Equation 51 of reference 1 is rewritten in the form of 12(A) or (B). In this equation the complement of the error function appears, which is a characteristic of diffusion.

$$t/\Theta=u \tag{10}$$

$$\zeta/\zeta_\infty=v \tag{11}$$

$$v=1-(1-\Phi\sqrt{u})\epsilon^u \tag{12A}$$

$$v=1-(\mathrm{erfc}\,\sqrt{u})\exp u \tag{12B}$$

The numerical calculation of equation 12 is easy for small values of u because both functions are well tabulated. For large arguments u, numerical values are hard to obtain, because it is necessary to handle a great number of decimals to obtain only a few significant figures. Errors are common because the expression erfc $\sqrt{u}$ decreases to zero whereas the expression exp u becomes infinite.

Another approach for large arguments u, is given by equation 13 which contains an asymptotic series. It is valid for u larger than 7.5 with an error of less than 0.1%.

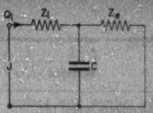
$$1-v=(\mathrm{erfc}\,\sqrt{u})\exp u=\frac{1}{\sqrt{\pi u}}\left[1-\frac{1}{2u}+\frac{3}{4u^2}-\frac{15}{8u^3}+\frac{105}{32u^4}\pm\ldots\right] \tag{13}$$

Numerical work is greatly simplified by using Fig. 5 and Table I which cover a wide range of values of the variables u and v. Beyond the given ranges the equations follow the indicated asymptotes. For low values of u this is equal to the plane approximation ($v=2\sqrt{u/\pi}$); for high values of u the series in equation 13 is unity ($1-v=1/\sqrt{u\pi}$).

Fig. 5 shows u and $1/u$ on a logarithmic scale on the abscissa, v and $1-v$ on another logarithmic scale on the ordinate, wherein v is a function of u and $1-v$ of $1/u$. Table I gives some values of v over a very wide range of u.

Long Pulses

Fig. 7. Equivalent circuit for heat dissipation for long pulses



The assumptions which led to the previous analysis were mainly that the semiconductor device contains continuously distributed heat-absorbing masses and thermal impedance arranged in a geometrical pattern. It was also assumed that no external heat-dissipation problems would arise during the short pulse duration of heat applied to the device. In reality, the diffusion wave reaches the boundaries of the device. If no other heat dissipation is provided, prolonged application of heat causes a temperature increase of the whole device. The increase of temperature of the surface over the ambient temperature causes heat dissipation to the atmosphere. At this stage these assumptions cease to be valid and another approximation must be used. The theory of continuous bodies becomes rather unwieldy and impractical if the problems of heat dissipation are also considered. One advantage of this theory which considers only mass and impedance is its flexibility, the ability to use it for almost any geometrical shape. By having to consider a third effect, this simplicity is lost because heat dissipation is not an equally distributed property of the materials. The difficulty is not as serious as it seems because it is only encountered for heat pulses which are relatively long, where the complexities of heat-wave propagation within the body of the semiconductor device lose their importance. The heat-diffusion wave, having expanded far enough to reach the outer boundaries of the device, leaves the inner sections at a temperature which is close to stationary heating and will be heated henceforth as one block.

Remaining in the realm of practical devices shows that the usual distribution of thermal masses lends itself to lumping of components. This is shown in Fig. 6(A) as a plot of incremental thermal impedance versus distance. The left-hand side of the figure represents the junction and the right-hand side represents the ambience. Fig. 6(B), on the same horizontal scale, shows the distribution of heat-absorbing masses. These figures show that a major lump of thermal impedance occurs near the junction and again between the cooling surface of the device and the ambient space. Inversely, the mass is lumped in the interval, where the incremental thermal impedance is low. By setting an arbitrary boundary B_i between the zone of predominant impedance and the zone of predominant

Table II. Material Constants for Metals

Material	k, W/M C	c, WS/KG C	$10^{-3}\gamma$, KG/M³	$10^{-6}\gamma c$, WS/M³ C	$10^{3}p$, M/$\sqrt{S}$	$10^{5}\tau$, M² C/W $\sqrt{S}$
Aluminum (pure) (Al)	204	920	2.68	2.46	9.11	4.47
Antimony (Sb)	18.3	165	6.68	1.10	4.07	22.5
Brass (70 Cu, 30 Zn)	120	380	8.10	3.08	6.25	5.20
Copper (pure) (Cu)	388	393	8.90	3 5	10.5	2.71
Germanium (Ge)	50	310	5.32	1.65	5.50	11.0
Indium (In)	24	238	7.25	1.725	3.73	15.5
Molybdenum (pure) (Mo)	146	270	10.20	2.76	7.28	4.99
Nickel (99.4 Ni+Co) (Ni)	60.7	470	8.66	4.07	3.86	6.37
Lead (99 Pb, 1 Sb) (Pb)	33	130	11.30	1.47	4.74	14.35
Silicon (Si)	84	735	2.34	1.72	7.00	8.31
Silver (pure) (Ag)	420	234	10.50	2.46	13.1	3.11
Carbon (C) steel (1.22 C, 0.35 manganese)	45	485	7.80	3.78	3.45	7.66
Steel (stainless) (no. *303*)	16.4	500	7.90	3.95	2.04	12.45
Tin (pure) (Sn)	65.6	235	7.40	1.74	6.15	9.35
Zinc (pure) (Zn)	110	400	7.20	2.88	6.18	5.61
Lead-tin (50 Pb, 50 Sn)	46.4	182	9.35	1.70	5.22	11.25
Soft solder (63 Sn, 37 Pb)	50.5	196	8.84	1.73	5.40	10.7
Zinc, commercial, rolled (0.08 Pb)	107	400	7.20	2.88	6.10	5.70
Gold (pure) (Au)	297	130	19.30	2.51	10.9	3.66
Tungsten (W)	187	142	19.3	2.74	7.81	4.67
Tantalum (Ta)	54.4	150	16.6	2.49	4.66	8.61
Columbium (Nb)	54	272	8.57	2.33	4.82	8.90

Thermal conductivity: k
Specific heat (mass): c
Specific mass: γ
Specific heat (volume): γc
Propagation constant: $p=(k/\gamma c)^{1/2}$
Temperature-rise constant: $\tau=(k\gamma c)^{-1/2}$

mass and another boundary B_c between the zones of predominant mass and of external heat dissipation, a lumped schematic diagram is obtained, Fig. 7. The internal thermal impedance Z_i represents the total sum of thermal impedances between the junction and the main mass of the device, whereas the external impedance Z_o represents the remainder of all thermal impedances counted from the main part of the mass to ambient space. The mass C is the sum of the product of the masses of the device times their specific heat.

The system is now subjected to a steady quantity of heat influx designated as Q_i flowing into the device through the internal thermal impedance Z_i.

Designate as θ the temperature rise of the junction (indicated by J) which follows equation 14:

$$CZ_o\theta'+\theta=Q_i(Z_i+Z_o) \qquad (14)$$

Solving for the temperature rise, and calculating as ζ the ratio of this temperature rise over the heat introduced at the junction, equation 15 is obtained:

$$\zeta_x=Z_i+Z_o[1-\exp(-t_x/CZ_o)] \qquad (15)$$

Correct initial conditions for this equation require matching to the prior analysis sections, such as for:

$$t_x=0,\ \zeta_x=Z_i,\ Z_i\approx Z_w \qquad (16)$$

and $t=t_x+T_w$

where the subscript x refers to the long pulse approximation, considered as another "section," and defined in the prior analysis. The subscript w refers to the last section analyzed before the lumped approximation is used. The time t is the running time measured on the same time base as for the other sections.

Graphical representation of the approximation follows the curves of the prior approximations without major deviation. If the assumptions are correct and within the properties of the device, the solution obtained by the two approximations should overlap in the interim range where the heat diffusion reaches the cooling surface. The time which appears in equation 15 is the same as the running time t of Fig. 4.

A substantial discrepancy between the results obtained by the two assumptions shows that excessive simplifications were made and a more detailed analysis is required. In this case it is preferable to work backwards by expanding the representation by lumped components to include more detailed impedance and heat-storing capacities. Since this approach involves the well-known analysis by a succession of *R-C* (resistance-capacitance) circuits, it is not included in this paper.

Complete Device

For numerical application of this method, Table II gives a summary of material constants. The treatment of a complete device is shown in detail in Appendix II.

Fig. 8 shows the cross section of a silicon diode, without enclosure or anode connector, together with the transient thermal impedance versus time curve for cathode cooling only. The numerical calculations for Fig. 8 are from Appendix II. This graphic representation shows how the successive approximations complement each other and that this method is flexible enough to cover very wide ranges of time and thermal impedances.

This method seems to be corroborated by the available test results. It is particularly helpful in guiding short-circuit or pulse load tests, in the interpretation of test curves, and in the analysis of failures. If the analysis must be undertaken with numerically inaccurate assumptions, the shape of the resulting curve helps to coordinate it with test results, which may be used to check back on the validity of the assumption.

Vol.
Pt.
196

Appendix I. Definitions of Spherical Approximation

The definitions of a spherical sector and its limits of thermal impedance and time are combined in Fig. 9. Fig. 9(A) shows the spherical section of a homogeneous

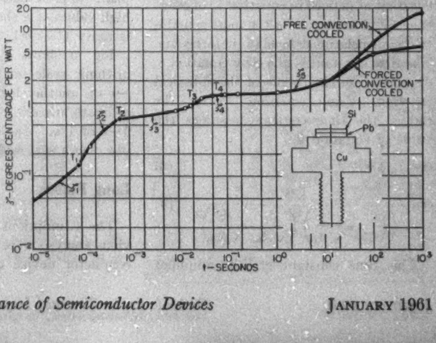


Fig. 8. Transient thermal impedance versus time for silicon diode analyzed in Appendix II

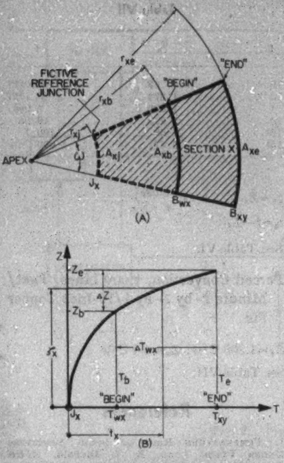


Fig. 9. Spherical section of homogeneous body (A) and its thermal impedances and times (B)

body. Section X refers to any section of the body which may be similar to Fig. 1. The spherical sector has an opening angle which is defined as a solid angle ω (and not a plane angle as shown in the figure). The solid angle is defined by the area A and the radius r of any spherical area which is cut out by the spherical sector. The solid angle is determined by equation 17. With this definition the stationary temperature rise follows equation 18.

$$\omega = A/r^2 \tag{17}$$

$$d\theta = -\left(\frac{q}{k\omega}\right)\frac{dr}{r^2} \tag{18}$$

Integrated within the boundaries of the BEGIN and the END of the section gives equation 19. The subscript x, pertaining to section X is omitted here. It is understood that these equations have different constants for each of the sections of the body.

$$\theta_b - \theta_e = \frac{q}{k\omega}\left(\frac{1}{r_b} - \frac{1}{r_e}\right) \tag{19}$$

$$\Delta Z = \Delta\theta/q = \frac{\dfrac{1}{r_b} - \dfrac{1}{r_e}}{k\omega} \tag{20}$$

Equation 19 gives also thermal impedance, as equation 20, in which ΔZ is the incremental thermal impedance of the section, between the BEGIN and the END as defined in Fig. 9(A). This incremental thermal impedance is also shown in Fig. 9(B) together with the time increment for this section.

Since the areas of the BEGIN and the END are known, equation 17 inserted in equation 20 yields the value for the solid angle, equation 21. This yields equation 22(A) and (B) for the radii shown in Fig. 9(A).

$$\omega = \left[\frac{\dfrac{1}{\sqrt{A_b}} - \dfrac{1}{\sqrt{A_e}}}{k\Delta Z}\right]^2 \tag{21}$$

$$r_b = \frac{\sqrt{A_b}\,k\Delta Z}{\dfrac{1}{\sqrt{A_b}} - \dfrac{1}{\sqrt{A_e}}} \tag{22A}$$

$$r_e = \frac{\sqrt{A_e}\,k\Delta Z}{\dfrac{1}{\sqrt{A_b}} - \dfrac{1}{\sqrt{A_e}}} \tag{22B}$$

Any particular analysis can only be fitted into the over-all analysis if the stationary thermal impedance Z_b at the BEGIN of the section is made equal to the stationary thermal impedance at the END of the previous section, indicated by Z_e in Fig. 9(B), i.e., stationary thermal impedances of successive sections must always be equal when referred to the same boundary. These boundaries are indicated in Fig. 9(B) by the letter B with the subscripts referring to the previous section w, section X under consideration, and the subsequent section y. The thermal impedance between the fictive junction J_x and the beginning B_{wx} of the section must be equal to Z_b. Equation 23 shows the value of this impedance which, when combined with equation 20, gives equation 24 for the area and equation 25 for the radius of the corresponding fictive reference junction of the section X.

$$Z_b = \frac{\dfrac{1}{\sqrt{A_j}} - \dfrac{1}{\sqrt{A_b}}}{\sqrt{\omega}\,k} \tag{23}$$

$$A_j = \left[\frac{\Delta Z}{\dfrac{Z_e}{\sqrt{A_b}} - \dfrac{Z_b}{\sqrt{A_e}}}\right]^2 \tag{24}$$

$$r_j = \frac{k\Delta Z^2}{\left(\dfrac{Z_e}{\sqrt{A_b}} - \dfrac{Z_b}{\sqrt{A_e}}\right)\left(\dfrac{1}{\sqrt{A_b}} - \dfrac{1}{\sqrt{A_e}}\right)} \tag{25}$$

These equations replace any section of a real device by a spherical section and determine a thermal impedance, characterized by a fictive junction which is remote from the section, a BEGIN area, END area, and a solid angle which numerically describes the lateral boundary (without requiring a more definite geometrical shape).

Appendix II. Example

As an example, see the device shown in Fig. 8 with the curve of transient thermal impedance versus time which is calculated in the following (see Table III).

Transient Thermal Impedance from Junction to Base

Plane Approximation.

1. Silicon (diffused junction):

$$\zeta_1 = \frac{2\tau_1\sqrt{t_1}}{A_1\sqrt{\pi}} = 14.55\sqrt{t_1}\ \text{C/w}$$

$$T_1 = \left(\frac{Z_1 A_1\sqrt{\pi}}{2\tau_1}\right)^2 = 9.12(10^{-5})\ \text{second}$$

$t_1 = t$	ζ_1
10^{-8}	$1.455(10^{-3})$
$9.12(10^{-5})$	0.139

2. Solder, $k = 33$ w/mC, $\gamma c = 1.47 \times 10^6$ ws/m^{3}C:

Table III. Stationary Thermal Impedance from Junction to Base

	Material	d, Meters	A, Meters^2	ΔZ, C/w	Z, C/w
1	Silicon	$7.5\ (10^{-5})$	$6.45\ (10^{-6}$	0.139	0.139
2	Solder	10^{-4}	$6.45\ (10^{-6})$	0.470	0.609
3	Copper	10^{-3}	$7.00\ (10^{-6})$	0.376	0.985
4	Copper	$3.05\ (10^{-3})$	$2.84\ (10^{-5})$	0.283	1.268

Table IV

t_3	u	v	ζ_3	t
$0.0119 = T_{32}$	0.0125	0.114	$0.609 = Z_2$	$6.505\ (10^{-4}) = T_2$
0.020	0.0210	0.144	0.766	$8.75\ (10^{-3})$
0.025	0.0262	0.158	0.840	$1.375\ (10^{-2})$
0.030	0.0314	0.171	0.910	$1.875\ (10^{-2})$
$0.0352 = T_{33}$	0.037	0.185	$0.985 = Z_3$	$2.395\ (10^{-2}) = T_3$

Table V

t_4	u	1/u	v	1−v	ζ_4	t
$3.12\ (10^{-3}) = T_{43}$	3.12	0.32	0.720	0.280	$0.985 = Z_3$	$2.395\ (10^{-2}) = T_3$
10^{-2}	10	0.10	0.830	0.170	1.14	$3.083\ (10^{-2})$
3×10^{-2}	30	0.033	0.90	0.100	1.24	$5.083\ (10^{-2})$
$5.89\ (10^{-2}) = T_{44}$	58.9	0.017	0.927	0.073	$1.268 = Z_4$	$7.973\ (10^{-2}) = T_4$

$$\zeta_2=\frac{2\tau_2\sqrt{t_2}}{A_2\sqrt{\pi}}=25\sqrt{t_2},\ \text{C/w}$$

$$T_{21}=\left(Z_1\frac{A_1\sqrt{\pi}}{2\tau_2}\right)^2=3.07(10^{-5}),\ \text{seconds}$$

$$T_{23}=\left(Z_2\frac{A_2\sqrt{\pi}}{2\tau_2}\right)^2=5.90(10^{-4}),\ \text{seconds}$$

$$t=t_2+T_1-T_{21};\quad T_2=T_{23}+T_1-T_{21}$$

t_2	ζ_2	t
3.07 (10^{-5})	0.139	9.12 $(10^{-5})=T_1$
10^{-4}	0.250	1.605 (10^{-4})
5.90 (10^{-4})	0.609	6.505 $(10^{-4})=T_2$

Spherical Approximation.

3. Copper:

$A_b=6.45\ (10^{-5})\ \text{m}^2;\ A_e=7.65\ (10^{-4})\ \text{m}^2;$
$Z_b=0.609\ \text{C/w};\ Z_e=0.985\ \text{C/w}$

From equation 12:

$\zeta_{\infty 3}=5.32\ \text{C/w}$

From equation 14:

$\Theta_3=0.954$ second

See Table IV.

$$t=t_3+T_2-T_{32};\quad T_3=T_{34}+T_2-T_{32}$$

4. Copper:

$A_b=7.65\ (10^{-4})\ \text{m}^2;\ A_e=9.3\ (10^{-5})\ \text{m}^2;$
$Z_b=0.985\ \text{C/w};\ Z_e=1.268\ \text{C/w}$

From equation 12:

$\zeta_{\infty 4}=1.377\ \text{C/w}$

From equation 14:

$\Theta_4=1.00\ (10^{-3})$ second

See Table V.

$$t=t_4+T_3-T_{43};\quad T_4=T_{45}+T_3-T_{43}$$

Thermal impedance of contact between base and fin, 1.7 C/w. Free convection, 2 by 2 by 1/16-inch copper fin.

$Z_t=1.268\ \text{C/w};\ Z_0=15.5\ \text{C/w}$

For diode and fin; $C=16.2$ ws/C. From equation 15:

Table VI

t_5	ζ_5	t
10^{-2}	1.268	0.0897
10^{-1}	1.274	0.1797
1	1.330	1.080
10	1.878	10.08
10^2	6.37	100.1
10^4	16.47	1,000

Table VII

t_5	ζ_5	t
10^{-2}	1.268	0.0897
10^{-1}	1.274	0.1797
1	1.330	1.080
10	1.873	10.08
10^2	4.58	100.1
10^3	5.67	1,000

$$\zeta_5=Z_t+Z_0(1-\epsilon^{-t_5/CZ_0})$$

$$t_5=t-T_4$$

See Table VI.

Forced Convection, 1,000 Linear Feet/Minute 2- by 2- by 1/16-Inch Copper Fin

$Z_t=1.268\ \text{C/w};\ Z_0=4.4\ \text{C/w}$

See Table VII.

References

1. Temperature Rise of Solid Junctions Under Pulse Load, E. J. Diebold. *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 76, Nov. 1957, pp. 593–98.

2. Tables of Probability Functions (book). "Mathematical Tables," Vol. I, U. S. Bureau of Standards, Washington, D. C., 1941.

Hybrid Computers for Process Control

GEORGE BIRKEL, JR.
NONMEMBER AIEE

THERE is a need in industry for equipment of moderate complexity and cost capable of solving simple process control equations at high speed. Quite economical process control and processing of instrumentation data may be accomplished by a hybrid computer. This tool combines the best features of analog and digital systems with a simplicity and economy unavailable in either, to produce a unique combination which is capable of operating at extremely high speeds and accuracies, over large numerical ranges.

The Radiation Incorporated hybrid analog-digital (AD) computing technique is based upon AD conversion devices used as ratiometers and digital-to-analog (DA) conversion devices used as multipliers. Addition and subtraction are accomplished by digital logic and summation. Exponent bookkeeping retains the full accuracy of the data, with automatic adjustment of scale factors. Round-off of data is accomplished in the sense of "best fit."

Hybrid Computing

Process control systems often require some form of AD and DA conversion. DA conversion (decoding) provides an analog output signal directly proportional to the digital input and to the analog reference signal used; Fig. 1. That is, the output is directly proportional to the product of the digital input and the analog reference. The input multiplicand, X, is referred to the standard analog reference, R, or "unity." The multiplier, Y, is referred to the multiplicand

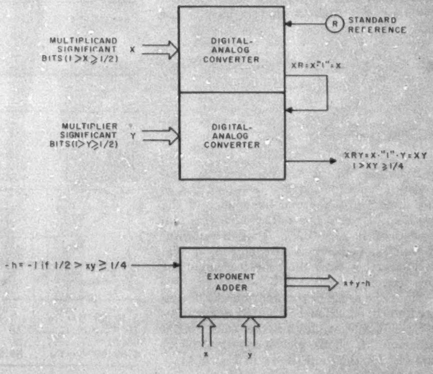


Fig. 1. Hybrid multiplication

Paper 60-978, recommended by the AIEE Computing Devices Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE-AIChE-IRE-ISA-ASME Joint Automatic Control Conference, Cambridge, Mass., September 6–9, 1960. Manuscript submitted May 24, 1960; made available for printing July 8, 1960.

George Birkel, Jr., is with Radiation Incorporated, Melbourne, Fla.

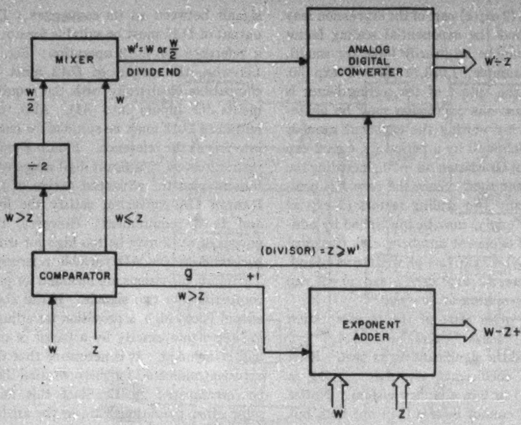

MIXER
W' = W or W/2
DIVIDEND
ANALOG DIGITAL CONVERTER
W÷Z
W/2
W
÷2
W>Z
W≤Z
(DIVISOR) = Z ≥ W'
COMPARATOR
g
+1
W>Z
EXPONENT ADDER
W-Z+g
W
Z
W
Z

Fig. 2. Hybrid division

analog-conversion level as a reference. The result, $XRY=X\cdot 1\cdot Y=XY$, is an analog signal representing a product.

The DA converter is used as the multiplier for the hybrid computer because of its logical simplicity and speed. Multiplication requires one "bit" time, using this method.

The input to the DA converter consists only of the significant digits of a number properly normalized in standard floating-point format. Therefore,

$$2^{-1}\leqslant X<1,\ 2^{-1}\leqslant Y<1$$

The exponents x and y of multiplicand and multiplier are added together to obtain the conditional product exponent. From the foregoing,

$$2^{-1}\leqslant XY<1$$

After normalizing, the final product exponent will be $(x+y-h)$, where $h=0$ or 1.

The digital output of an AD converter is directly proportional to its analog input, W, and inversely proportional to its reference, Z. Therefore, the AD converter performs the equivalent of the division $W\div Z$; Fig. 2. The input $W\leqslant Z$ for proper operation of the equipment. If $W>Z$, W is attenuated by the fixed factor 2. When attenuation is required, the quotient exponent must be incremented. The dividend for a division may be the output of a DA converter or some data analog level. The divisor is the analog equivalent of the digital divisor required, referred to R. The exponents w and z, corresponding to the dividend and divisor, provide the quotient exponent $(w-z+g)$, where $g=0$ or 1 is determined by the attenuation condition. Division requires one "word" time.

A digital adder-subtractor is used. The AD converter operates serially, high-bit-first, to perform division. The output product-quotient has the exponent $(x+y+g-z-h)$, as noted before. Exponents x, y, and z are known at time of DA entries. Serial addition of significant bits of a term, 2^kK, may be started at time $(k-(x+y-z))$, with respect to AD

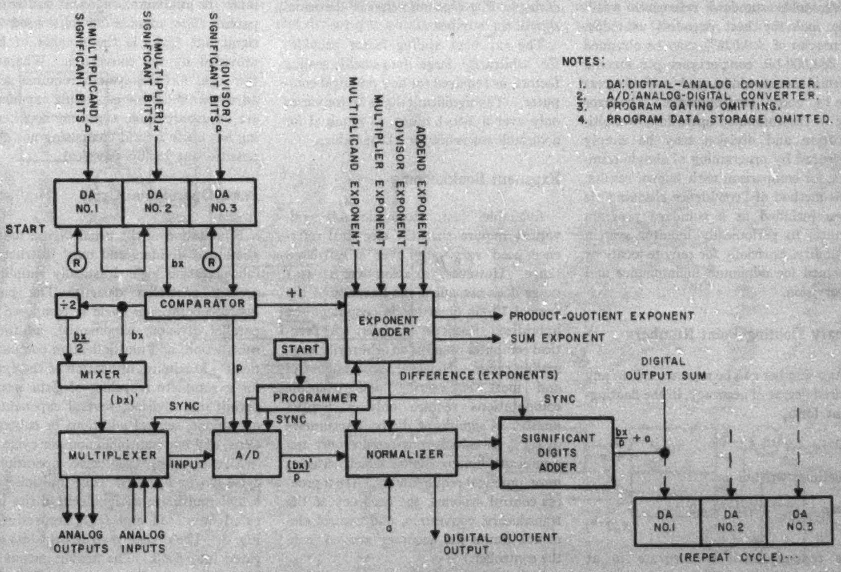

NOTES:
1. DA: DIGITAL-ANALOG CONVERTER.
2. A/D: ANALOG-DIGITAL CONVERTER.
3. PROGRAM GATING OMITTING.
4. PROGRAM DATA STORAGE OMITTED.
(MULTIPLICAND) SIGNIFICANT BITS b
(MULTIPLIER) SIGNIFICANT BITS x
(DIVISOR) SIGNIFICANT BITS a
MULTIPLICAND EXPONENT
MULTIPLIER EXPONENT
DIVISOR EXPONENT
ADDEND EXPONENT
DA NO.1
DA NO.2
DA NO.3
START
R
bx
R
÷2
COMPARATOR
+1
EXPONENT ADDER
PRODUCT-QUOTIENT EXPONENT
SUM EXPONENT
bx/2
bx
START
p
MIXER
DIFFERENCE (EXPONENTS)
DIGITAL OUTPUT SUM
(bx)'
PROGRAMMER
SYNC
SYNC
SYNC
MULTIPLEXER
INPUT
A/D
(bx)'/p
NORMALIZER
SIGNIFICANT DIGITS ADDER
bx/p + a
a
ANALOG OUTPUTS
ANALOG INPUTS
DIGITAL QUOTIENT OUTPUT
DA NO.1
DA NO.2
DA NO.3
(REPEAT CYCLE)···

Fig. 3. Basic operational cycle of hybrid computing system

high-bit output, to adjust for exponent disparity. By delaying adder operation by one bit time, the effect of $g=1$ is provided. If $h=1$, the high bit of the product-quotient is 0, but no add-time compensation is needed. Parallel addition may be preferred.

Utilizing the AD converter for division and the DA converter for multiplication drastically reduces arithmetic costs, since the converters are usually in the process control system anyway. Moreover, the simplicity of hybrid multiplication and division ordinarily means less programing. Utilizing the programing inherent in multiplexing of raw data is straightforward for a hybrid system. (Programed multiplexing pioneered by Radiation Incorporated is a standard feature for the RADIPLEX units.) Where multiplication and division predominate, hybrid programing is especially simple. Most mathematical expressions of physical and engineering principles predominate in multiplication and division terms, with relatively few additions or subtractions. Such conversions may usually be performed with sufficient speed for time-sharing of raw data conversions and arithmetic operations.

Multiplication and division accuracies are functions of the converter quantizing errors. Round-off therefore, is taken in the sense of "best fit." Using the most stable standard references available and the best encoders available, accuracies of ±0.025% may be obtained at 384,000-bit comparisons per second. Normal care should be used to avoid errors due to "noisy" environments. The error due to the combined operations of multiplication and division may be simply measured by programing of simple numbers, for comparison with known results. This method of "confidence checking" is often included as a standard program feature, to periodically monitor system reliability, especially for remote locations designed for minimum maintenance and supervision.

Binary Floating-Point Numbers

Any number can be represented, to any desired degree of accuracy, in the floating-point form,

$$2^x(2^{-1}+A_2 2^{-2}+A_3 2^{-3}+\ldots+A_n{}^{x-n})$$

sometimes written as

$$(2 \exp x)\,(2^{-1}+A_2 2^{-2}+A2^{-3}+\ldots+A_n 2^{-n})$$

The representation is accurate to at least 1 part in 2^n or better.

The (2 exp x) part of the expression may be called the exponential scaling factor and may be arbitrarily large (or small). For example, 1,073,741,824 = 2 exp 30. Since the base 2 of the scaling factor is constant, the expression may be abbreviated by writing the exponent number only, followed by a period (.), e.g., 2 exp 30 may be written as +30., including the exponent sign. Since the base 2 is common, any two scaling factors (2 exp x) and (2 exp y) may be multiplied by addition of exponent numbers, e.g., (2 exp x) (2 exp y) = 2 exp $(x+y)$, written as $(x+y)$. Likewise, (2 exp x) ÷ (2 exp y) = 2 exp $(x-y)$, written as $(x-y)$.

The other part of the representation $X=(2^{-1}+A_2 2^{-2}+A_3 2^{-3}+\ldots+A_n 2^{-n})$, is called the significant digits part. Each of the coefficients, A_2, A_3, …, A_n, is either 1 or 0 in a binary system. Notice that X cannot be less than one half, but must be less than one.

Any number will be written in floating-point form, as:

exponent sign	exponent, base 2	·	significant digits = fraction	algebraic sign

The exponential scaling factor, including sign, is written as the binary equivalent of the exponent only and determines the relative size or scale of a number. The significant digits as a fraction $(<1, \geqslant 2^{-1})$ identify a number to the specified accuracy. The algebraic sign is the usual sign of the number.

The exponent scaling factor provides for arbitrarily large (or small) scaling factors, as required for any practical computer. The significant digits factor varies only over a 2-to-1 range, as required for a variable reference for AD encoding.

Exponent Bookkeeping

Available, fast, electronic AD converters require that the electrical reference used vary only over a restricted range. However, variation over a 2-to-1 range does not affect the linearity of the normal 12-bit binary-code output of the hybridized RADICON encoder. A practical computer requires an arbitrarily large scaling range. However, process control and most engineering and scientific computations require only a limited number of significant digits. Accuracies of 0.1% in calculations require only ten significant binary digits, which satisfies most practical requirements. (In a process control systems, the accuracy of the transducers, converters, and control elements, limits the accuracy needed from the controller.)

A hybrid system requires control of signals between all its converters. The output of $DA1$ must be suitable for use as a reference for $DA2$ operation; Fig. 3. Likewise, the output of $DA2$ must be compatible electrically with the requirements for inputs into AD. Also, the output of $DA3$ must be suitable for entry into AD as the reference. Floating-point restrictions on significant digit range variations, plus the advanced design of the RADIAN DA converter, satisfy the first and third requirement. However, the output of $DA2$ may be too high for comparison with the AD variable reference. Therefore, provisions are included for precomparing the two signals. If the $DA2$ output is too high, a provision for adjusting the signal exactly by a factor of one half is included. It is necessary that the exponent indicated for this operation then be incremented by 1. Since this (and some other operations) are in the analog domain, it may not properly be called a floating-point operation. To signify both digital and analog exponent scaling operations, the expression "exponent bookkeeping" is used.

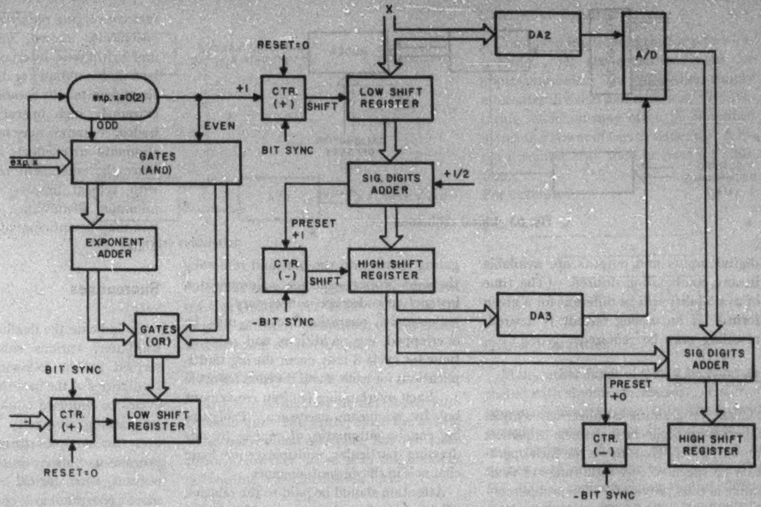
The number of bits used for the exponent field depends upon the maximum (or minimum) exponent size anticipated. The number of bits used for significant digits is the number of bits provided by the converters. Wherever the logic for the system requires scale adjustment, the corresponding exponents are adjusted; thus, exponent bookkeeping has made hybrid computing not only possible but highly practical.

Basic Operational Cycle

Elimination of the usual digital multipliers and dividers and most instruction interpretation logic drastically simplifies process controller designs. The basic hybrid operational cycle is composed of parallel division, exponential, addition, subtraction, and multiplication computations. Maximum utilization of the cycle for a complete recycling of data would permit one division, several exponential operations, several additions or subtractions, and one multiplication per cycle.

Algebraically, the basic operational cycle is $((bx \div p)+a)$. The multiplicand b and multiplier x significant digits are gated into $DA1$ and $DA2$, respectively, Fig. 3. The divisor p significant bits are gated into $DA3$. The analog output of $DA2$ is compared to output of $DA3$, for

Fig. 4. Square root


x
DA2
A/D
RESET=0
+1
exp. x≡0(2)
CTR. (+)
SHIFT
LOW SHIFT REGISTER
ODD
EVEN
exp. x
GATES (AND)
BIT SYNC
SIG. DIGITS ADDER
+1/2
PRESET +1
EXPONENT ADDER
CTR. (-)
SHIFT
HIGH SHIFT REGISTER
-BIT SYNC
DA3
GATES (OR)
SIG. DIGITS ADDER
PRESET +0
BIT SYNC
CTR. (+)
LOW SHIFT REGISTER
CTR. (−)
HIGH SHIFT REGISTER
RESET=0
-BIT SYNC

range compatibility. If *DA*2 output exceeds *DA*3 output, *DA*2 output is attenuated by 2, incrementing exponent for product-quotient by 1. Modified *DA*2 output is sampled as input to AD and *DA*3 output is used as AD reference, i.e., divisor. Digital output of AD becomes the addend into the adder, being added to the augend, *a*. (Other additions could have been performed during these operations, resulting in *a*.) The adder output is then ready as an input into the next cycle, in digital form.

Likewise, the operational cycle $((b \div q)+a)y)$ starts with the analog dividend *b* and divisor *q* to obtain a digital quotient from the AD converter. The quotient is then added to *a*. This sum is next multiplied by *y*, obtaining an analog product. The cycle may then be repeated.

Each of these cycles is a different form of the other and may be continuously repeated. Obviously, both analog and

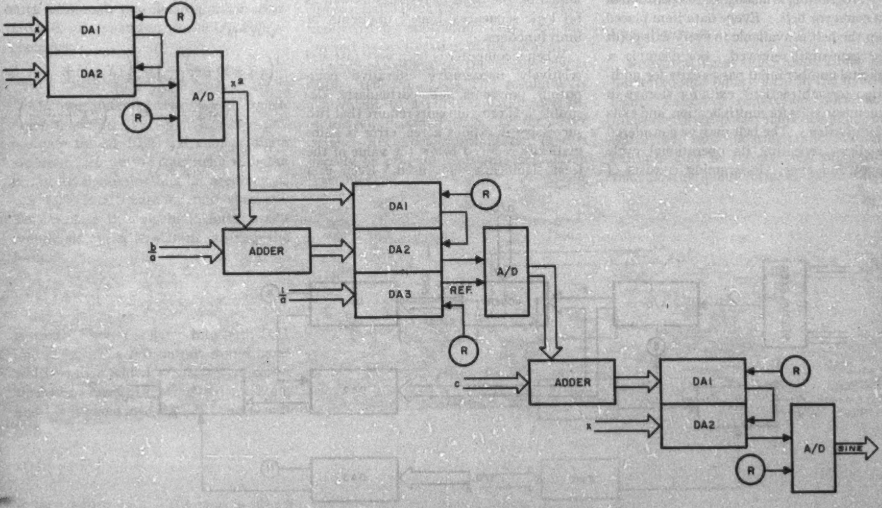

x
DA1
R
x
DA2
A/D
R
x²
b/a
ADDER
DA1
R
DA2
A/D
1/a
DA3
REF.
R
c
ADDER
DA1
R
x
DA2
A/D
SINE
R

Fig. 5. Calculation of sine approximation $(\sin\ (\pi/2x = ax^5+bx^3+cx)$

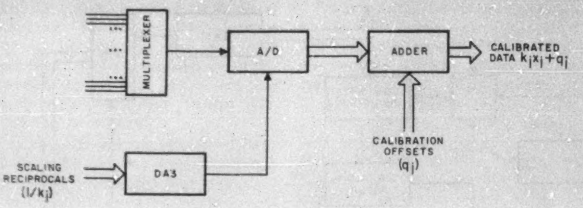

MULTIPLEXER
A/D
ADDER
CALIBRATED DATA $k_jx_j+q_j$
CALIBRATION OFFSETS (q_j)
SCALING RECIPROCALS $(1/k_j)$
DA3

Fig. 6. Linear calibration

digital inputs and outputs are available in any combination desired. (The time of availability will be different for a given form.) If an analog output is desired, rescaling may be required.

Programing

Cyclic programing is inherently simple. Operations occur in a definite sequence, requiring only that inputs for each operation be specified. A wide range of flexibility is thus provided. For process control and other applications where a wired-in program could be utilized for long periods of time, the sequence of data items may also be wired in. However, if random-data sequence is desirable, the data may be so entered. Usual memory media are used, although the operational cycle removes the need for storing some dynamic results which are in circulation instead. For some systems it is simpler to resample inputs than to store them.

Programing is analogous to operation of a conveyor belt. Every data item placed on the belt is available to every exit point or gate until removed. For example, a digital number input passes exits for addition (or subtraction), exits for storage in memory, exits for multiplication, and exits for division. The belt may be considered endless, repeating its operational cycle over and over. Programing consists of gating items onto the belt and removing them in proper sequence. Complicated instruction codes are unnecessary.

Successive operational cycles may be overlapped, e.g., additions and subtractions for cycle 3 may occur during multiplications for cycle 2 and division for cycle 1. Such overlapping is often convenient but by no means necessary. Programing can be automatic, of course, by addressing particular multiplexer or logic channels in the desired sequence.

Attention should be paid to the relative efficiencies of various approaches. Approximation techniques which depend on iterative improvements or convergence are ideal. Thiele's continued-fraction expansions are quite useful in hybrid programs in contrast to digital programs, where division is avoided when computing time must be minimized. Standard double or triple precision techniques, corresponding to the use of partial fractions, may be utilized. Branch points may be determined on the basis of residues as well as bit logic sequences, binary quotients, or limit functions.

When a computer is both very fast and relatively inexpensive, iterative computing programs are particularly desirable. If program exits require that two successive iterations agree, error is automatically reduced below the value of the least significant digit used. Even if a fast converging recurrence relation is not specifically known, some sort of logic and arithmetic interpolation test might be found to obtain iterative routines. For these reasons, the combined economy and extremely high operating speed of the hydrid computer may be used to virtually eliminate error effects in such programs. Since the equipment used is compatible with typical process control elements, minimum disturbance of normal operation may be anticipated for most applications.

Subroutines

To indicate the flexibility of the hybrid computer, various subroutines are described in the following sections. The similarities of the basic hybrid operational cycles to standard equations indicate the programing simplicity available. However, for complex programs, arithmetic program simplicity is relatively less important than logical simplicity. Since most programing logic functions are digital for large, complex programs, the programing system required is that of a general-purpose digital computer. A subroutine should be designed to utilize the full hybrid operational cycle wherever feasible.

SQUARE ROOT, FIG. 4

Since

$$(X^2 \div (X\pm\Delta)) = X \mp \Delta + \Delta^2/(X\pm\Delta)$$

a better approximation than $(X\pm\Delta)$ to $\sqrt{X^2}$ is

$$\frac{1}{2}\left((X\pm\Delta)+\frac{X^2}{X\pm\Delta}\right)=\frac{1}{2}\left((X\pm\Delta)+(X\mp\Delta)+\frac{\Delta^2}{X\pm\Delta}\right)=\frac{1}{2}\left(2X+\frac{\Delta^2}{X\pm\Delta}\right)=X+\frac{\Delta^2}{2(X\pm\Delta)}$$

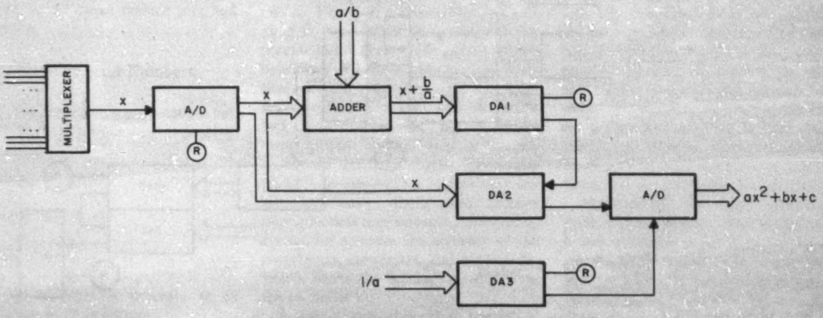


Fig. 7. Thermocouple (quadratic) linearization

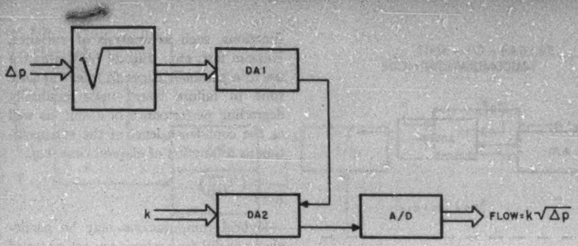


Fig. 8. Flow calculation

provided that

$$\Delta > \frac{\Delta^2}{2(X\pm\Delta)}$$

i.e., if

$$X > -\frac{\Delta}{2}; \text{ or } X > \frac{3\Delta}{2}$$

which is simply assured.

A simple principle permits obtaining $(X\pm\Delta)$, where $\Delta \leqslant X/40$, in a single addition and logic step from X^2. Since the square root of the exponent is exactly known, the principle is applied only to the significant digits. The argument term, X^2, is gated into $DA1$, "1/2" is gated into $DA2$, and the first approximation $(X\pm\Delta)$ is gated into $DA3$. Since the output of $DA2$ is the AD input and the output of $DA3$ is the AD reference, the coder digital output is the quotient, $1/2\ (X^2 \div (X\pm\Delta))$. When added to $1/2\ (X\pm\Delta)$, X is known accurately to about 1/3,120. Since the error is always positive, by subtracting 1/6,240, X is known to ±1/6,240, if necessary.

SINE FUNCTION, FIG. 5

A power-series approximation to sin $(\pi/2)\,x$, where $-1 \leqslant x \leqslant 1$, is $cx+bx^3+ax^5$, accurate to ±1/10,000, if enough digits are used. However, consistent with the use of 12 significant bits an accuracy of ±1/4,096 is adequate for most purposes. To calculate this algebraic form would require six cycles as shown; instead, the form

$$y=x\left(c+\frac{x^2}{1/a}\left(\frac{b}{a}+x^2\right)\right)$$

is used. First, x is gated into both $DA1$ and $DA2$. The AD output would be x^2 which is then added to b/a. The sum is then gated into $DA2$, x^2 is gated into $DA1$, and $1/a$ is gated into $DA3$. The AD output

$$a(x^2)\left(\frac{b}{a}+x^2\right)$$

is then added to c. The sum is placed in $DA1$, x is gated into $DA2$. The AD output is then the digital approximation to sin $(\pi/2)x$, ±1/4,096. Cos $(\pi/2)x$ may be similarly computed, or calculated as a function of sin $(\pi/2)x$.

POWER SERIES (GENERAL)

A power series may be developed term by term quite efficiently, since the hybrid basic operational sequence has the same form. However, as seen in the previous subroutine, the calculation form

$$\left(\left(\frac{xy}{q}+t\right)\frac{z}{p}+\ldots\right)$$

may be even more efficient.

RECIPROCAL FUNCTIONS

Sometimes it is desirable to multiply by the reciprocal of a number rather than use division, e.g., for an analog output preceded by "division." Reciprocal multiplication may be simply accomplished by using certain simple linear approximations of the form $y=(A-bx)$, computed by subtraction only, and improved by an iteration of the form $y(2-yx)=z$. As indicated, the product xy^2 is subtracted from $2y$. The resultant is equivalent to a cubic equation, but requires fewer computations than the power series. Reciprocal multiplication is used mostly on program constants, seldom on variables.

CONTINUED FRACTIONS

Continued fractions of the form $(a+b\div(c+d\div(\ldots)))$ are sometimes used for approximations. The calculation simply alternates division and addition. By factoring, addition may often be simplified by using a power of two as addend. A binary counter may then be used for addition.

For example:

$$e^x=1+\cfrac{x}{1-\cfrac{x}{2+\cfrac{x}{3\left(1-\frac{1}{3}\,\frac{(x)}{2+\ldots}\right)}}}$$

OTHER FUNCTIONS

Many other functions may be approximated with similar efficiencies. The best approximations for use with the hybrid computer predominate in multiplication and division. Finite difference techniques used are similar to those required for digital computers, or digital difference analyzers, except that a successive sampling may be used to obtain differences. Differential and integral equations are usually approximated by standard asymptotic, hypergeometric, and power-series expressions. For example, for small x,

$$J_P(x) \sim \frac{1}{2^P P!}\,x^P$$

is frequently used in solving second-order differential equations.

DATA FORMATS

Numerical bases other than 2 may be converted into binary form in the hybrid using standard conversion formulas. Likewise, fixed-point numbers may be converted into floating-point format. Outputs may be reconverted into other than floating-point formats, as required.

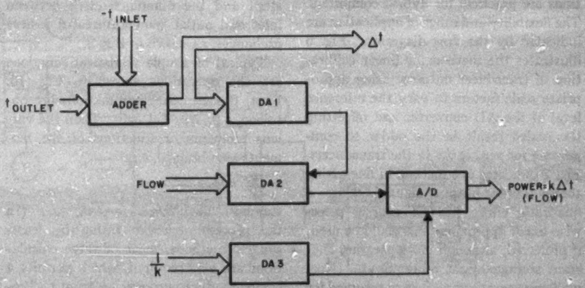


Fig. 9. Power calculation

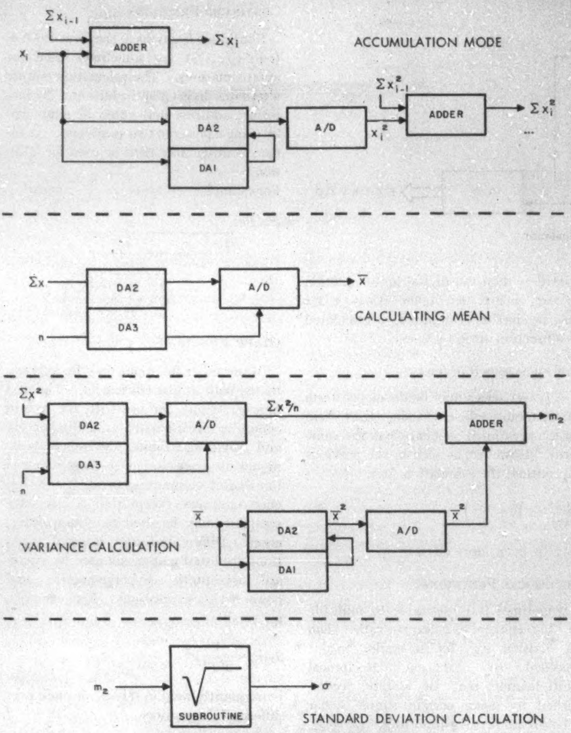


Fig. 10. Simple statistical calculations

The relative economy of multiplying and dividing may be utilized heavily in such conversions.

Applications

A wide range of process control operations are practical for hybrid computers. The flexibility and range of application are indicated by the flow diagrams. Fig. 6 illustrates the method for linear calibration of transducer outputs using appropriate scale factors to vary the reference level of the AD converter and offsetting the scaled result in the adder to compensate for zero shifts in the transducers. Fig. 7 indicates one method of linearizing thermocouple output signals, using a quadratic curve fit. Of course, piecewise linear approximations could be used, if preferred, although more program constant storage might be required. (Normalizing operations necessary assumed for all figures, although not specifically shown.) All basic units are used over and over, as in a digital computer.

Fluid flow through a system is sometimes calculated simply by an expression of the form $k\sqrt{\Delta p}$, as shown in Fig. 8. Power output is often calculated as the product of flow and some program constant and the difference (ΔT) between inlet and outlet temperatures of a heat exchanger, as shown in Fig. 9.

Typical of simple statistical functions are the operations shown in Fig. 10. First, the necessary summations are obtained, e.g., Σx, Σx^2, etc. Then the various moments or functions of the moments are obtained, e.g.,

$$\bar{x} = \Sigma x/n, m_2 = \Sigma x^2/n - \bar{x}^2$$

standard deviation $\sigma = \sqrt{m_2}$, ... (In the process of curve fitting by least-squares methods, most of these calculations are also required, which permits a compact dual-purpose program.) Programs are available for more complex functions, such as analysis of variance, Duncan test, etc. Fig. 11 illustrates the use of a parabolic curve fit to extrapolate time to failure based upon gradually degrading performance of a unit, as well as the confidence level of the extrapolation as a function of elapsed time, e.g.,

$$R = e^{-\left(\frac{t_f - t_i}{t_f}\right)}$$

Hybrid computations may be particularly useful for preliminary estimates of material requirements. As a somewhat oversimplified example, consider Fig. 12 which illustrates the estimation of the weight (*kg*) of fissionable nuclear material required for a 2-group 4-region cylindrical critical reactor. The weight in kilograms is given by:

$$\text{Weight} = x(\pi R^2)\rho H$$

where

x = volumetric fraction of fissionable material
R = core radius
ρ = density constant for pure fission material
H = reactor height

The volumetric fraction, x, of fissionable material, or the reactor dimensions, R and H, are assumed to have been previously estimated to a satisfactory accuracy, probably by an iteration between upper and lower limits based upon radial buckling, macroscopic cross sections, and reflector properties.

System Compatibility

Because of its converters, a hybrid computer is compatible with a wide range of analog and digital devices. As indicated under "Applications," conversion of digital formats is feasible. An unusual feature is the possibility for hybrid operations with a normalized "Grey" or permuted binary code. By use of odd-even parity information, a Grey-code hybrid system is feasible.

Hybrid computers will probably be used most widely for supplementing more expensive or slower systems, doing simple routine calculations more efficiently than other machines. For example, by linearizing data to calibrate transducer system outputs, the hybrid computer frees more expensive or slower digital machines for more complex calculations which require general-purpose or universal logic capabilities.

The recycling nature of hybrid operations and the availability of both analog and digital inputs and exists, give the hybrid computer an iterative or feedback characteristic, highly compatible with process control systems.

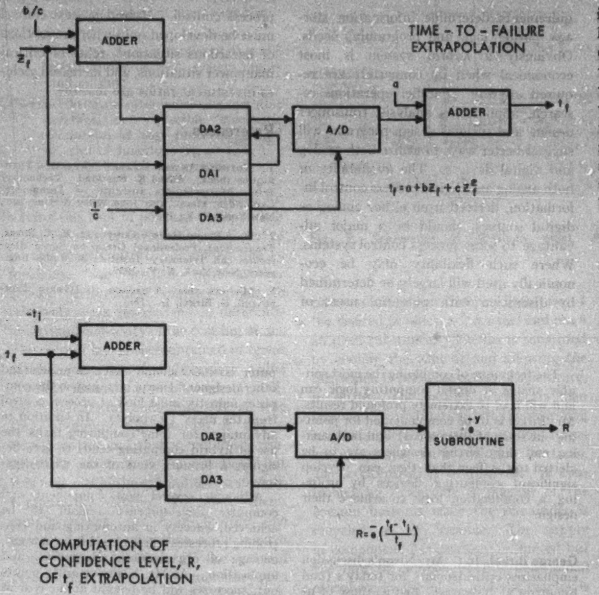

b/c
ADDER
z_f
TIME - TO - FAILURE EXTRAPOLATION
a
ADDER
t_f
DA2
A/D
DA1
$\frac{1}{c}$
DA3
$t_f = a + bz_f + cz_f^2$
$-t_i$
ADDER
t_f
DA2
A/D
-y
SUBROUTINE
R
DA3
$R = e^{-\left(\frac{t_f - t_i}{t_f}\right)}$
COMPUTATION OF CONFIDENCE LEVEL, R, OF t_f EXTRAPOLATION

Fig. 11. Extrapolating operational time to failure

Conclusions

Advances in analog-digital-analog conversion devices led to the development of a hybrid (analog and digital) computer. Floating-point multiplication and division can be accomplished in bit and word conversion times, respectively. Many iterations, interpolations, series, and asymptotic approximations can be accomplished in two or three operational cycles using conventional AD conversion units.

The hybrid computer will be used largely for linear calibration and linearization in the immediate future, but will find restricted use in nuclear process monitoring and alarm systems, simple batch control, and other simple but very repetitive operations. A novel use may be automatic checking of system functions. By observing the effect of placing an analog component between the DA and AD converters, a trouble-shooting routine over a range of conditions is obtained. By varying the input digital patterns into the DA units, the variations in the AD output will provide a close look into analog unit variations under dynamic conditions. Likewise, the perturbations due to process parameter variations might be examined, perhaps with some measure of optimization control. As with all interdisciplinary engineering, the computer designer must learn more about users' needs before the real utility of hybrid computing will be realized.

Hybrid computation techniques are adaptable to the solution of many simple problems of process control and computation where on-line computing is desired. These techniques combine the performance of digital and analog equipment to produce a unique combination capable of operating at extremely high speeds and at high accuracies. The low cost, small size, and low power consumption permit on-line, economical use of hybrid computers at remote locations.

The accuracy of a hybrid computer is defined by the accuracies of available converters and system logic extensions. Partial-fraction techniques may be used for any desired degree of accuracy at the cost of increased logic complexity. Analog control units are more restricted in accuracies. In the sense that an analog computer or any analog device could be inserted between the DA units and the AD unit, and any digital device including a computer, can be inserted between the AD and DA units, the hybrid computer represents most data acquisition and data-reduction systems and therefore has extremely universal characteristics. The systems described in this paper are concerned with the basic hybrid computer requirements.

In terms of immediate applications, the hybrid systems described will supplement or replace analog equipment when automatic setup capability or increased accuracies are required. Likewise, expen-

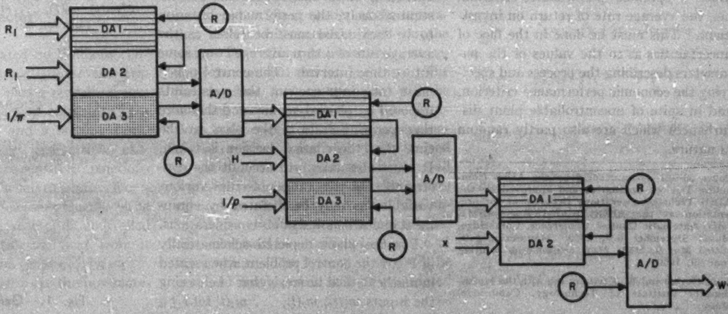


Fig. 12. Computation of $WT(K_e) = x \times (R_1^2/1/\pi)(H/1/\rho)$

sive digital computers will be relieved of routine linearizing and editing functions. The basic capability of a hybrid system is simple arithmetic transformation during its conversion from analog to digital or from digital to analog domains. If transformation from either domain to the other is needed in a system, then hybrid computing may be feasible.

The speed, accuracy, and flexibility requirements in an application establish basic hybrid system costs. A basic system requires one AD and three DA units, one significant digits adder, one exponent adder, one high-shift normalizer and one low-shift normalizer. Other system requirements determine information storage, gating, logic and programing needs. Obviously a hybrid system is most economical when its converters are required anyway. Parallel operations research, applications analysis, transducer design, and systems design programs will suggest better ways to utilize both analog and digital devices. The availability of both analog and digital process control information, derived from either analog or digital sources, should be a major advantage to some process control systems. Where such flexibility may be economically used will largely be determined by discussion with potential users of process control. Hybrid process control must be developed only where reductions of hazardous situations, relief of critical manpower situations, and increased yield-to-investment ratios are feasible.

References

1. Notes on Analog-Digital Conversion Techniques (book), **Alfred K. Susskind.** Technology Press, Massachusetts Institute of Technology, Cambridge, Mass.; also John Wiley & Sons, Inc., New York, N. Y., 1957.

2. High Speed Hybrid Computer, **M. C. Burns.** *Transactions*, Professional Group on Space Electronics and Telemetry, Institute of Radio Engineers, New York, N. Y., 1959.

3. Mathematical Approach to Hybrid Computing, **G. Birkel, Jr.** *Ibid.*

◆

Discussion

Floyd E. Nixon (The Martin Company, Orlando, Fla.): It is always difficult to tell how profound something is when you do not know your position in the forest. For example, techniques of computing which are very conventional and easy in digital computing will not appear profound to a digital man; but to accomplish these same mathematical objectives may be very difficult if analog computing techniques are used. Therefore, utilizing digital techniques might be extremely profound to an analog man. The same is true for reverse cases.

The technique of combining the most suitable analog or digital computing logic can and will lead to extremely profound results. Mr. Birkel is to be complimented for pointing out some of these areas; and both analog and digit circuit designers are to be alerted to the fact that they can develop significant computing devices by utilizing a combination logic to achieve their design.

George Birkel, Jr.: Mr. Nixon's discussion emphasizes critical points for today's (and tomorrow's) designers. Just as those of us who represent various aspects of the computer business should strive to understand other designers' viewpoints, just so the computer industry must look at process control from the users' viewpoints. In addition to advantages for some computing tasks, the use of hybrid computing tends to give designers a broader view of the whole system.

Although several new component and computer developments remain to be achieved, success in automating and controlling processes will lead us into a new era, an age of organization, adaptation, and application. As Mr. Nixon so aptly points out, successes will be bought at the cost of interdisciplinary understanding.

Dynamic Optimization and Control of a Stirred-Tank Chemical Reactor

L. A. GOULD
ASSOCIATE MEMBER AIEE

W. KIPINIAK
NONMEMBER AIEE

THE CONTROL of a chemical plant may be viewed as an attempt to vary the plant inputs in such a way as to maximize a specified performance criterion, i.e., the average rate of return on investment. This must be done in the face of uncertainties as to the values of the parameters describing the process and specifying the economic performance criterion, and in spite of uncontrollable plant disturbances which are also partly random in nature.

Since plant conditions and disturbances forever vary with time, and because the plant itself is a dynamic system whose operating state cannot be changed instantaneously, the performance criterion to be maximized must be taken as the average rate of return, averaged over some future time interval. The control policy must take into account the constraints imposed by plant dynamics and therefore the operating state of the plant at the time that the control decision is taken. It must also take into account the constraints on the range of the various variables as may be dictated by equipment limitations and safety considerations.

For the plant depicted schematically in Fig. 1 the control problem, when stated in mathematical terms, is that of choosing the inputs $m_1(t)$, $m_2(t)$, $\ldots$ $m_l(t)$ for $t \geq t_0$ to maximize the expected average rate of return $\mathcal{P}$, where

$$\mathcal{P}=E\left\{\frac{1}{t_e-t_0}\int_{t_0}^{t_e}[\mathcal{S}(q_1(t),q_2(t),\ldots q_k(t))-\mathcal{C}(m_1(t),m_2(t),\ldots m_l(t))]\,dt\right\} \qquad (1)$$

where

t_0 = present time, i.e., the time at which the control decision is taken
t_e-t_0 = time interval chosen for averaging rate of return
$\mathcal{S}[q_1(t), q_2(t), \ldots q_k(t)]$ = sales price of plant outputs $q_1, q_2, \ldots q_k$
$\mathcal{C}[m_1(t), m_2(t), \ldots m_l(t)]$ = cost of plant inputs $m_1, m_2, \ldots m_l$

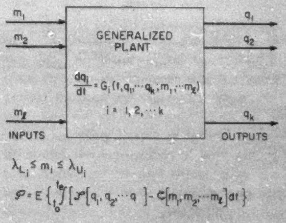

m1
m2
mℓ
INPUTS
GENERALIZED PLANT
dqi/dt = Gi(t, q1, ⋯ qk; m1, ⋯ mℓ)
i = 1, 2, ⋯ k
q1
q2
qk
OUTPUTS
λLi ≤ mi ≤ λUi
𝒫 = E{∫t0te [𝒮[q1, q2, ⋯ q] − 𝒞[m1, m2, ⋯ mℓ] dt}

Fig. 1. Generalized plant

Paper 60-982, recommended by the AIEE Computing Devices Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE-AIChE-IRE-ISA-ASME Joint Automatic Control Conference, Cambridge, Mass., September 6–9, 1960. Manuscript submitted May 6, 1960; made available for printing June 29, 1960.

L. A. Gould and W. Kipiniak are with the Massachusetts Institute of Technology, Cambridge, Mass.

$\lambda_{Li} \leq m_i \leq \lambda_{Ui}$, $i=1, 2, \ldots l$, specify constraints on range of input variables

$dq_i(t)/dt = G_i(t;\ q_1, q_2, \ldots q_k;\ m_1, m_2, \ldots m_l)$, $i=1, 2, \ldots k$, represent dynamic relationships between system inputs and outputs

E = expectation operator indicating ensemble average over possible stochastic variations of cost functions $\mathcal{S}$ and $\mathcal{C}$, and of functions G_i, $i=1, 2, \ldots k$, which specify the system

Of the many methods available for solving this type of problem the approach of the calculus of variations is believed to be the most fruitful. It allows one to formulate the control problem for general performance criteria, subject to relatively weak restrictions on differentiability and concavity of the cost functions. The control law is then obtained in terms of the solutions of a boundary value problem associated with a set of ordinary differential equations. It is true that obtaining the solution is not a trivial matter and that it requires a considerable amount of computation, but it is well within the capabilities of modern computating devices. In many chemical processes, for example the one described below, the time scale is such that on-line solution of the control problem is feasible and the monetary gain is such as to justify considerable investment in the more sophisticated control equipment needed to perform such calculations.

It is not the purpose of this paper to discuss the variational method of control system synthesis in general; this has been done elsewhere.[1-3] Rather, an attempt will be made to show how this method is applied to a specific problem in chemical process control, namely, the control of the stirred-tank reactor postulated by Williams and Otto in their excellent paper.[4] This reactor is believed to represent a typical chemical process and the problems encountered in controlling it, typical of chemical process control in general.

The Williams Reactor

The plant postulated by Williams consists of a stirred-tank reactor fed two pure input streams plus a recycle stream and a separation section consisting of a distillation column and a settling tank. In the reactor three irreversible exothermic temperature-sensitive reactions take place to yield a mixture of six components in the reactor effluent. The temperature in the reactor is controlled by adding or subtracting heat from it through heating and cooling coils.

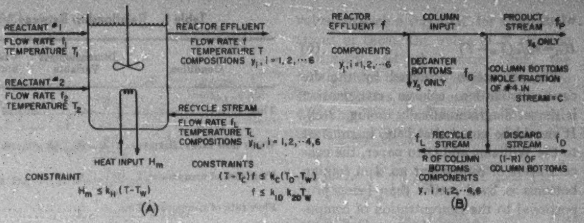


Fig. 2. Schematics of Williams plant. A—Reactor section. B—Separation section

Components 1 and 2 are the raw materials, no. 3 is an intermediate, no. 4 the desired product, no. 5 a tar, and no. 6 an inert substance. The tar is separated by cooling the effluent and allowing the tar, which is insoluble at low temperature, to settle. The desired product is then removed by distillation but the stripping is not complete because the product forms an azeotrope with column bottoms. The concentration of the inert is controlled by discarding a portion of the column bottoms while the remainder is recycled to the reactor. The supply of component 2 is assumed unlimited but the flow of component 1 is determined elsewhere in the plant and the reactor is required to process all that it receives.

The reactor through-put is limited by column capacity, which in turn depends on cooling water temperature. Reactor temperature and through-put are also limited by the ability of the cooler to reduce effluent temperature to the point where the tar will precipitate in the decanter. The rate of heat supply to the reactor, or its removal, is limited by the capacity of the reactor heating and cooling coils.

The above is an abridged description of the plant postulated by Williams. Reference is made to his paper for a more thorough description of the plant and for a derivation of the mathematical description of its behavior.[4]

In this study only the reactor will be considered. The operation of the separation section will manifest itself on the reactor only through limitations on reactor through-put and temperature. The recycle flow and composition will be assumed to be deterministic time functions independent of short-term fluctuations in reactor operation but constrained to satisfy, on the average, the over-all material balance. Thus, all the control settings in the recovery section, including the fraction of column bottoms being discarded, will be considered set at a constant value. The price structure will be assumed fixed and it will be postulated that the entire production of component 4 can be sold.

The reactor considered is shown schematically in Fig. 2(A) and the equations describing its behavior are listed in the following. The equations are all in normalized, dimensionless form and the relation between the normalized, dimensionless variables and those used by Williams is shown in Table I, together with the numerical values of the constants involved.

The reactions taking place are

$$
\begin{aligned}
&1(\text{no. }1)+1(\text{no. }2)\xrightarrow{r_1}2(\text{no. }3)\\
&1(\text{no. }2)+2(\text{no. }3)\xrightarrow{r_2}1(\text{no. }4)+2(\text{no. }6)\\
&2(\text{no. }3)+1(\text{no. }4)\xrightarrow{r_3}3(\text{no. }5)
\end{aligned}
\tag{2}
$$

The reaction rates are

$$
\begin{aligned}
r_1 &= y_1 y_2 e^{\alpha_1-\frac{\beta_1}{T}}\\
r_2 &= y_2 y_3 e^{\alpha_2-\frac{\beta_2}{T}}\\
r_3 &= y_3 y_4 e^{\alpha_3-\frac{\beta_3}{T}}
\end{aligned}
\tag{3}
$$

The material balance for each component is

$$
\begin{aligned}
\frac{dy_1}{dt} &= f_1+f_L y_{1L}-f y_1-r_1\\
\frac{dy_2}{dt} &= f_2+f_L y_{2L}-f y_2-r_1-r_2\\
\frac{dy_3}{dt} &= f_L y_{3L}-f y_3+2r_1-2r_2-2r_3\\
\frac{dy_4}{dt} &= f_L y_{4L}-f y_4+r_2-r_3\\
\frac{dy_5}{dt} &= -f y_5+3r_3\\
\frac{dy_6}{dt} &= f_L y_{6L}-f y_6+2r_2
\end{aligned}
\tag{4}
$$

The reactor heat balance is

$$
\frac{dT}{dt} = f_1 T_1+f_2 T_2+f_L T_L-fT+h_1 r_1+h_2 r_2+h_3 r_3+H_m \tag{5}
$$

where

$$H_m = k_H f_w (T_w - T) \qquad (6)$$

The separation achieved by the decanter-distillation column combination is shown diagrammatically in Fig. 2(B). It should be noted that here, in contrast to the Williams and Otto paper, the concentration of component no. 4 in column bottoms is fixed rather than being proportional to the concentration of component no. 6. Thus the various output streams are

$$f_P = f\left(\frac{y_4 - C(1-y_6)}{1-C}\right) \qquad (7)$$

$$f_G = f y_6 \qquad (8)$$

$$f_D = f(1-y_4-y_6)\left(\frac{1-R}{1-C}\right) \qquad (9)$$

The monetary rate of return is therefore

$$\mathcal{S} - \mathcal{C} = P_P f_P + P_D f_D + P_G f_G - P_A f_1 - P_B f_2 - P_U f - P_F \qquad (10)$$

In the above equations m(no. n) represents m units of component n; y_n = weight fraction of component no. n in the reactor, and y_{nL} = weight fraction in the recycle; r_i, $i = 1, 2, 3$ are the reaction rates, and h_i = corresponding amounts of heat liberated; f_1, f_2, and f_L = flow rates of components 1, 2, and of the recycle, and T_1, T_2, and T_L = the corresponding stream temperatures; f = effluent flow rate, and T = reactor temperature; H_m = heat input to the reactor; f_w = water or stream flow rates through the reactor cooling or heating coils, and T_w = corresponding temperature; f_P, f_G, and f_D = output flow rates of the product, the tar and the discard streams, and P_P, P_G, and P_D = sales values; P_A and P_B = costs of components no. 1 and no. 2, P_U = utility cost, and P_F = fixed plant charges; R = fraction of the column bottoms being recycled, and C = azeotrope point of no. 4 in column bottoms.

In order to achieve a satisfactory separation of component 5 in the decanter, the effluent temperature must be lowered to below a critical temperature T_c and $f(T-T_c)$ units of heat removed. The capacity of the effluent cooler can be expressed as $k_c(T_0 - T_w)$, where T_0 is the cooler outlet water temperature. This imposes a constraint

$$(T-T_c)f \leq k_c(T_0 - T_w) \qquad (11)$$

Similarly, column capacity can be represented as $k_{1D} - k_{2D}T_w$, which imposes a constraint

$$f \leq k_{1D} - k_{2D}\, T_w \qquad (12)$$

The reactor cooling-heating coils' capacity is represented as $k_H(T - T_w)$, which requires that

$$|f_w| \leq 1 \qquad (13)$$

Inequality constraints cannot be readily treated by the techniques of the calculus of variations, but their effect can be incorporated in the control law by including in the rate of return expression penalty functions which are analytic, and which subtract little from the return rate when the constraints are satisfied but a lot when they are exceeded. Ideally, these should represent the cast of the damage resulting from exceeding the set limits. In most cases, however, any convenient function which approximates this behavior can be used.

Thus in the following problem it is convenient to use a penalty function $F(x) = x^{16}$, whose argument is the ratio of the constrained variable to the constraint. The modified monetary rate of return is therefore written as

$$\mathcal{S} - \mathcal{C} - k_{p1} F\left(\frac{(T-T_c)f}{(T_0-T_w)k_c}\right) - k_{p2} F\left(\frac{f}{k_{1D} - k_{2D}\, T_w}\right) - k_{p3} F(f_w) \qquad (14)$$

and in this way the restrictions on the range of the system variables are effectively incorporated in the control problem.

The performance criterion to be maximized can now be written as

$$\mathcal{P} = \left\{\frac{1}{t_e - t_0}\int_{t_0}^{t_e}\left[P_P\left(\frac{y_4 - C(1-y_6)}{1-C}\right)\times f + P_D(1-y_4-y_6)\left(\frac{1-R}{1-C}\right)f + P_G y_6 f - P_A f_1 - P_B f_2 - P_U f - P_F - k_{p1}F\left(\frac{T-T_c)f}{(T_0-T_w)k_c}\right) - k_{p2} \times F\left(\frac{f}{k_{1D}-k_{2D}\, T_w}\right) - k_{p3}\, F(f_w)\right]dt\right\} \qquad (15)$$

The recycle flow rates and compositions

Table I. Relationship Between Variables: Dimensionless and Williams'

Condition	Dimensionless Variable	Williams' Variable	
Time	t	t hr	$t = \frac{\bar{F}_R}{V} t = 20.65\, t$
Water flow rate	f_w	F_w	$f_w = \frac{F_w}{F_{w\max}} = 9.6 \times 10^{-5} F_w$
Flow rate of component 1	f_1	F_A lb/hr	$f_1 = \frac{F_A}{\bar{F}_R} = 1.042 \times 10^{-5} F_A$
Flow rate of component 2	f_2	F_B lb/hr	$f_2 = \frac{F_B}{\bar{F}_R} = 1.042 \times 10^{-5} F_B$
Recycle flow rate	f_L	F_L lbs/hr	$f_L = \frac{F_L}{\bar{F}_R} = 1.042 \times 10^{-5} F_L$
Effluent flow rate	f	F_R lbs/hr	$f = \frac{F_R}{\bar{F}_R} = 1.042 \times 10^{-5} F_R$
Reactor temperature	T	T_R R*	$T = \frac{T_R}{\bar{\bar{T}}_R} = 1.5625 \times 10^{-3}\, T_R$
Recycle temperature	T_L	T_L R	$T_L = \frac{T_L}{\bar{\bar{T}}_R} = 1.5625 \times 10^{-3}\, T_L$
Component 1 inlet stream temperature	T_1	T_A R	$T_1 = \frac{T_A}{\bar{\bar{T}}_R} = 1.5625 \times 10^{-3}\, T_A$
Component 2 inlet stream temperature	T_2	T_B R	$T_2 = \frac{T_B}{\bar{\bar{T}}_R} = 1.5625 \times 10^{-3}\, T_B$
Reaction 1 rate constant	k_1	k_1 lb/lb, hr	$k_1 = \frac{V}{\bar{F}_R} k_1 = 0.0484\, k_1$
Reaction 2 rate constant	k_2	k_2 lb/lb, hr	$k_2 = \frac{V}{\bar{F}_R} k_2 = 0.0484\, k_2$
Reaction 3 rate constant	k_3	k_3 lb/lb, hr	$k_3 = \frac{V}{2\bar{F}_R} k_3 = 0.0242\, k_3$
Reaction 1 rate constant	α_1	A_1	$\alpha_1 = \ln(0.0484\, A_1) = 19.4833$
Reaction 2 rate constant	α_2	A_2	$\alpha_2 = \ln(0.0484\, A_2) = 25.5569$
Reaction 3 rate constant	α_3	A_3	$\alpha_3 = \ln(0.0242\, A_3) = 33.0827$
Reaction 1 activation parameter	β_1	B_1 R	$\beta_1 = \frac{B_1}{\bar{\bar{T}}_R} = 18.7500$
Reaction 2 activation parameter	β_2	B_2 R	$\beta_2 = \frac{B_2}{\bar{\bar{T}}_R} = 23.4370$
Reaction 3 activation parameter	β_3	B_3 R	$\beta_3 = \frac{B_3}{\bar{\bar{T}}_R} = 31.2500$
Heat of reaction 1	h_1	H_1 Btu/lb	$h_1 = \frac{2H_1}{\bar{\bar{T}}_R C_P} = 0.9771$

* Degrees Rankine.

Table I (cont.)

Condition	Dimensionless Variable	Williams' Variable	
Heat of reaction 2	h_2	H_2 Btu/lb	$h_2=\dfrac{3\,H_2}{\overline{T}_R\,C_P}=0.5860$
Heat of reaction 3	h_3	H_3 Btu/lb	$h_3=\dfrac{3H_3}{\overline{T}_R C_P}=1.6750$
Cooling water temperature	T_w	T_{wt} R	$T_w=\dfrac{T_{wt}}{\overline{T}_R}$
Effluent cooler capacity	k_c		$k_c=\dfrac{F_{LCW}C_W}{\bar{F}_R\,C_P}=4.000$
Reactor heat exchanger capacity k_H	k_H		$k_H=\dfrac{UA}{C_P\,\bar{F}_P}=0.2606$
Column capacity specification	k_{1D}		$k_{1D}=22.2$
Calumn capacity specification	k_{2D}		$k_{2D}=25.6$
Recycle ratio	R		$R=0.551$
Azeotrope point	C		$C=0.038$
Price of product	P_P	P_P	$P_P=\dfrac{P_P(1-0.124)\ 8{,}400\overline{F}_R}{2{,}780{,}000}=76.212$
Price of discard stream	P_D	P_D	$P_D=\dfrac{P_D(1-0.124)\ 8{,}400\ \bar{F}_R}{2{,}780{,}000}=1.716$
Price of tar	P_G	P_G	$P_G=\dfrac{-P_G\,8{,}400\ \bar{F}_R}{2{,}780{,}000}=-2.900$
Cost of component 1	P_A	P_A	$P_A=\dfrac{P_A\ 8{,}400\ \bar{F}_R}{2{,}780{,}000}=5.800$
Cost of component 2	P_B	P_B	$P_B=\dfrac{P_B\ 8{,}400\ \overline{F}_R}{2{,}780{,}000}=8.700$
Fixed charges	P_F	P_F	$P_F=\dfrac{P_F}{2{,}780{,}000}=0.099$
Utility costs	P_U	P_U	$P_U=\dfrac{P_U\ 8{,}400\ \bar{F}_R}{2{,}780{,}000}=0.076$
Mean reactor temperature	T_0	$\overline{T}_R=640$ R	
Mean effluent cooler outlet water temperature	T_0	$T_{w0}=540$ R	$T_0=0.844$
Critical decanter temperature	T_c	$T_c=580$ R	$T_c=0.906$
Mean reactor effluent flow rate		$\bar{F}_R=95{,}961$ lb/hr	
Effluent cooler water flow rate		$F_{ECW}=153{,}500$ lb/hr	
Reactants heat capacity		$C_p=0.40$ Btu/F, lb	
Water heat capacity		$C_w=1.00$ Btu/F, lb	
Reactor holdup		$V=4{,}640$ lb	
Reactor heat exchanger heat transfer coefficient		$U=100$ Btu/ft², F, hr	
Reactor heat exchanger effective surface area		$A=100$ ft²	

are treated as independent time functions, i.e., independent of reactor operation. They therefore have the status of disturbances. The function f_1 is also an independent time function determined elsewhere in the plant so that its variations are disturbances. Similarly, variations in the temperature of cooling water and feed streams are disturbances. It will be assumed, however, that they are all deterministic, so that the controller knows about them at least t_e-t_0 in advance. Under such circumstances the expectation operator can be dropped and the problem becomes purely deterministic at every time that a control decision is to be taken. (The design procedure being presented is capable of treating stochastic signals; this is one of its advantages. However, only the deterministic case will be treated here to keep the exposition simple.)

The system behavior for $t \geq t_0$ also depends on the current values of the state variables of the process. The state variables are those which cannot be altered instantaneously by any finite inputs and only their derivatives can change discontinuously. In this problem the reactor temperature T and the reactor compositions have this property, as shown in equations 4 and 5.

The control variables of the process are f_2 and f_w. The controller has to adjust them on the basis of its knowledge of the system state—values of $T(t_0)$ and $y_1(t_0)$, $\ldots y_5(t_0)$—and the values of disturbances for $t_0 \leq t \leq t_e$. The function which defines the value of $f_2(t_0)$ and $f_w(t_0)$ to be applied for the current system state and future disturbances is the control law. The object of the design procedure is to find the control law which maximizes $\mathcal{P}$.

Euler Equations for the Williams' Reactor

To maximize $\mathcal{P}$ as expressed in equation 15 by the variational calculus, one attempts to choose the time functions involved in such a way that, for small variations in them, the value of $\mathcal{P}$ remains stationary. At the same time, the relationships between the control and the state variables, as expressed by equations 4 and 5, must be maintained. To do this one forms a new performance criterion:

$$\begin{aligned}\mathcal{P}=\Bigg\{\frac{1}{t_e-t_0}\int_{t_0}^{t_e}\Bigg[\Bigg(&P_P\left(\frac{y_4-C(1-y_5)}{1-C}\right)+\\&P_D(1-y_4-y_5)\left(\frac{1-R}{1-C}\right)+\\&P_G y_5-P_U\Bigg)f-P_A f_1-P_B f_2-P_F-\\&k_{P1}F\left(\frac{(T-T_c)f}{(T_0-T_w)k_c}\right)-k_{P2}\times\\&F\left(\frac{f}{k_{1D}-k_{2D}\,T_w}\right)-k_{P3}F(f_w)+\\&\omega_1\,(-y_1'+f_1+f_L y_{1L}-fy_1-r_1)+\\&\omega_2\,(-y_2'+f_2+f_L y_{2L}-fy_2-r_1-r_2)+\\&\omega_3\,(-y_3'\quad+f_L y_{3L}-fy_3+2r_1-\\&\quad 2r_2-2r_3)+\\&\omega_4(-y_4'\quad+f_L y_{4L}-fy_4\ +r_2-r_3)+\\&\omega_5\,(-y_5'\qquad\quad -fy_5\quad +3r_3)+\\&\omega_6(-T'+f_1T_1+f_2T_2+f_LT_L)-\\&\quad fT+h_1\,r_1+h_2\,r_2+h_3\,r_3+k_Hf_w\times\\&\qquad\qquad (T_w-T)\Bigg]dt\Bigg\}\end{aligned} \quad (16)$$

and requires that $\mathcal{P}$ be stationary not only with respect to variations in $f_2, f_w, y_1, \ldots y_5, T$, but also with respect to variations in $\omega_1, \omega_2, \ldots \omega_6$. Clearly the only way $\mathcal{P}$ can be stationary with respect to variations in $\omega_1, \omega_2, \ldots \omega_6$ is for their coefficients to vanish, i.e., for the constraint equations to be satisfied. Furthermore, when the constraint equations are satisfied, $\mathcal{P}$ of equation 16 has the same value as $\mathcal{P}$ of equation 15. The variables $\omega_1, \omega_2, \ldots \omega_6$ so introduced are often called the Lagrange multipliers.[8]

Taking variations of $\mathcal{P}$ with respect to the time functions involved and equating them to zero, one obtains

$$\begin{aligned}\delta_{f_2}\mathcal{P}=0=-\omega_1y_1-\omega_2(y_2-1)-\omega_3y_3+&\\ y_4\left[P_P\frac{1}{1-C}-P_D\frac{1-R}{1-C}-\omega_4\right]+&\\ y_5\left[P_P\frac{C}{1-C}-P_D\frac{1-R}{1-C}+P_G-\omega_5\right]+&\\ \omega_6(T_2-T)-\Bigg[P_B+P_U+&\\ P_P\frac{C}{1-C}+P_D\frac{1-R}{1-C}\Bigg]-&\\ \frac{\partial}{\partial f}\left[k_{P1}\,F\left(\frac{(T-T_c)f}{(T_0-T_w)k_c}\right)\right]-&\\ \frac{\partial}{\partial f}\left[k_{P2}\,F\left(\frac{f}{k_{1D}-k_{2D}\,T_w}\right)\right]&\end{aligned} \quad (17)$$

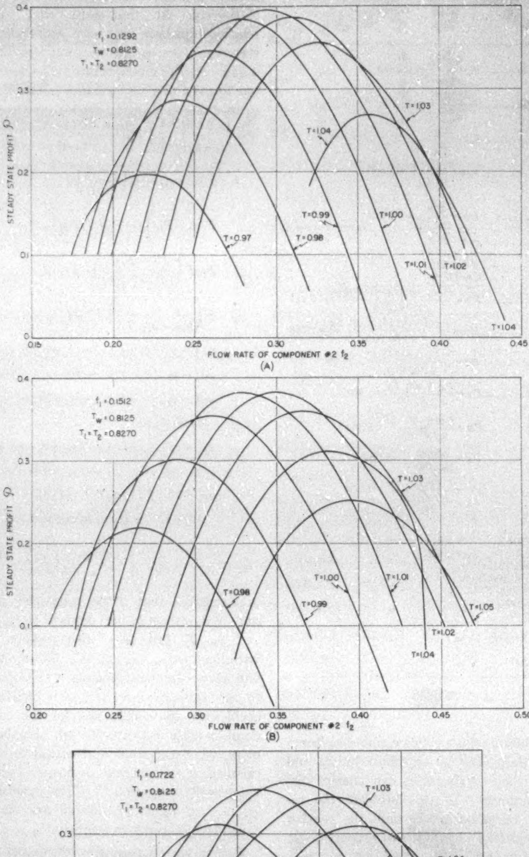


Fig. 3. Steady-state profit functions. A—At $f_1=0.1292$. B—At $f_1=0.1512$. C—At $f_1=0.1732$

$$\delta_{f_w}\mathcal{P}=0=\omega_6 k_H(T_w-T)-\frac{\partial}{\partial_{f_w}}[k_{P2}F(f_w)] \tag{18}$$

$$\delta_{y_1}\mathcal{P}=0=\omega_1'-\left(f+\frac{r_1}{y_1}\right)\omega_1-\frac{r_1}{y_1}\omega_2+\frac{2r_1}{y_1}\omega_3+\frac{h_1r_1}{y_1}\omega_6 \tag{19}$$

$$\delta_{y_2}\mathcal{P}=0=\omega_2'-\frac{r_1}{y_2}\omega_1-\left(f+\frac{r_1}{y_2}+\frac{r_2}{y_2}\right)\omega_2+2\left(\frac{r_1}{y_2}-\frac{r_2}{y_2}\right)\omega_3+\frac{r_2}{y_2}\omega_4+\left(\frac{h_1r_1}{y_2}+\frac{h_2r_2}{y_2}\right)\omega_6 \tag{20}$$

$$\delta_{y3}\mathcal{P}=0=\omega_3'-\frac{r_2}{y_3}\omega_2-\left(f+\frac{2r_3}{y_3}+\frac{2r_2}{y_3}\right)\omega_3+\left(\frac{r_2}{y_3}-\frac{r_3}{y_3}\right)\omega_4+\frac{3r_3}{y_3}\omega_5+\left(\frac{h_2r_2}{y_3}+\frac{h_3r_3}{y_3}\right)\omega_6 \tag{21}$$

$$\delta_{y4}\mathcal{P}=0=\omega_4'-\frac{2r_4}{y_4}\omega_3-\left(f+\frac{r_3}{y_4}\right)\omega_4+\frac{3r_3}{y_4}\omega_5+\frac{h_3r_3}{y_4}\omega_6+f\left(\frac{P_P}{1-C}-P_D\frac{1-R}{1-C}\right) \tag{22}$$

$$\delta_{y_5}\mathcal{P}=0=\omega_5'+f\left(P_P\frac{C}{1-C}-P_D\frac{1-R}{1-C}-P_G-\omega_5\right) \tag{23}$$

$$\delta_T\mathcal{P}=0=\omega_6'-\frac{r_1\beta_1}{T^2}\omega_1-\left(\frac{r_1\beta_1}{T^2}+\frac{r_2\beta_2}{T^2}\right)\omega_2-\left(\frac{2r_3\beta_3}{T^2}+\frac{2r_2\beta_2}{T^2}-\frac{2r_1\beta_1}{T^2}\right)\omega_3-\left(\frac{r_3\beta_3}{T^2}-\frac{r_2\beta_2}{T^2}\right)\omega_4+\frac{3r_3\beta_3}{T^2}\omega_5-\left(k_Hf_w-\frac{h_1\,r_1\,\beta_1}{T^2}-\frac{h_2\,r_2\,\beta_2}{T^2}-\frac{h_3\,r_3\,\beta_3}{T^2}+f\right)\omega_6-\frac{\partial}{\partial T}\left(k_{P1}F\left[\frac{(T-T_c)f}{(T_0-w)k_c}\right]\right) \tag{24}$$

$$\delta_{\omega_1}\mathcal{P}=0=-y_1'+f_1+f_Ly_{1L}-fy_1-r_1 \tag{25}$$

$$\delta_{\omega_2}\mathcal{P}=0=-y_2'+f_2+f_Ly_{2L}-f\,y_2-r_1-r_2 \tag{26}$$

$$\delta_{\omega_3}\mathcal{P}=0=-y_3'+f_Ly_{3L}-fy_3+2r_1-2r_2-2r_3 \tag{27}$$

$$\delta_{\omega_4}\mathcal{P}=0=-y_4'+f_Ly_{4L}-fy_4+r_2-r_3 \tag{28}$$

$$\delta_{\omega_5}\mathcal{P}=0=-y_5'\quad-fy_5\quad+3r_3 \tag{29}$$

$$\delta_{\omega_6}\mathcal{P}=0=-T'+k_Hf_w(T_w-T)+h_1r_1+h_2r_2+h_3r_3+f_1T_1+f_2T_2+f_LT_L-fT \tag{30}$$

where x' denotes dx/dt.

If one uses $F(x)=x^{16}$ as the form of the penalty function, equations 17 and 18 can be solved for f_2 and f_w, to yield

$$f_2=-f_1-f_L+\Bigg\{\Bigg[-\omega_1 y_1-\omega_2(y_2-1)-\omega_3 y_3+y_4\left(P_P\left(\frac{1}{1-C}\right)-P_D\left(\frac{1-R}{1-C}\right)-\omega_4\right)+y_5\left(P_P\left(\frac{C}{1-C}\right)-P_D\left(\frac{1-R}{1-C}\right)+P_G-\omega_5\right)-\omega_6(T_2-T)-\left(P_B+P_U+P_P\left(\frac{C}{1-C}\right)+P_D\left(\frac{1-R}{1-C}\right)\right)\Bigg]\Bigg/\left[16\,k_{P1}\left(\frac{(T-T_c)}{(T-T_w)k_c}\right)^{16}+16\,k_{P2}\left(\frac{1}{k_{1D}-k_{2D}T_w}\right)^{16}\right]\Bigg\}^{1/15} \quad (31)$$

$$f_w=\left\{\frac{\omega_6 k_H(T_w-T)}{16k_{P2}}\right\}^{1/15} \quad (32)$$

When these are substituted into equations 19 through 30 there results a set of 12 ordinary first-order differential equations in the state variables $y_1, y_2, \ldots y_5, T$, and the Lagrange multipliers $\omega_1, \omega_2, \ldots \omega_6$. These are the Euler equations for this problem.

The Euler equations are collectively of twelfth order and so 12 boundary conditions are needed to specify the solutions. The solutions represent the set of functions $y_1, y_2, \ldots y_5, T$, which for a specified change in operating state—from $y_1(t_0), y_2(t_0), \ldots y_5(t_0), T(t_0)$ to $y_1(t_e), y_2(t_e), \ldots y_5(t_e), T(t_e)$—maximize $\mathcal{P}$. At the time the control decision is taken, the present state of the system specifies $y_1(t_0), y_2(t_0), \ldots y_5(t_0), T(t_0)$. The conditions at $t=t_e$ and, for that matter, the value of t_e remain to be specified. Since the time over which the rate of return is to be averaged is probably many times the characteristic time constant of the system, t_e may be considered to approach infinity. If the expected values of the disturbances can be considered to approach constant values, the optimum solutions of the Euler equations must approach the steady-state optimum. Thus, at $t=t_e$, $y_1, y_2, \ldots y_5, T$ must all equal the values which yield a maximum of $\mathcal{P}$ in the steady state, and t_e must be large with respect to the characteristic time constants of the system.[2,3]

The boundary conditions on the solution of the Euler equations are specified at two points in time. Therefore, one must solve a 2-point boundary value problem. The values of the control efforts f_2 and f_w depend on $\omega_i(t_0)$, $i=2, \ldots 6$, and thus depend on the solution of the 2-point boundary value problem. Other ways of formulating the problem result in different specifications on $y_1, y_2, \ldots y_5, T$ at $t=t_e$, but in every case there results a 2-point boundary value problem. In every case, therefore, the control law is a function specified by the solution of a boundary value problem and the controller becomes a boundary-value problem solver.

Steady-State Optimization

Before solving the Euler equations and calculating the optimum control law, it was necessary to establish the desired values of the state variables at $t=t_e$, i.e., the values which result in maximum steady-state $\mathcal{P}$. For this purpose a computer program was written to map out the steady-state profit function for various operating conditions and to determine the optimum setting of the control variables f_2 and f_w for specified values of f_1 and cooling water temperature T_w.

To satisfy the steady-state material balance, the recycle flow rate and composition were made functions of reactor effluent flow rate and composition. Reference to Fig. 2(B) shows these relations to be

$$f_L=(1-y_4-y_5)\left(\frac{R}{1-C}\right)f \quad (33)$$

$$f_L y_{iL}=Ry_i f,\ i=1,2,3,\text{ and }6 \quad (34)$$

$$f_L\, y_{4L}=(1-y_4-y_5)\left(\frac{CR}{1-C}\right)f \quad (35)$$

and, since $f=f_1+f_2+f_L$, then

$$f=\frac{f_1+f_2}{1-(1-y_4-y_5)\left(\dfrac{R}{1-C}\right)} \quad (36)$$

In this way the over-all material balance is satisfied and variations in f_L considered in the dynamic optimization will be required to have an average value satisfying equation 33.

Examination of equation 5 indicates that the reactor itself is dynamically unstable. Therefore, it was not possible to simulate the reactor equations to compute $\mathcal{P}$ for various values of f_w. Instead, equation 4 was solved assuming T fixed, $\mathcal{P}$ was evaluated, and then f_w was chosen so as to satisfy equation 5 in the steady state.

Therefore the conditions to be satisfied in the steady state were

$$\begin{aligned}
&f_1-y_1(1-R)f && -r_1 &&=0\\
&f_2-y_2(1-R)f && -r_1-r_2 &&=0\\
&-y_3(1-R)f && +2r_1-2r_2-2r_3 &&=0\\
&-\left(y_4-(1-y_4-y_5)\ \frac{CR}{1-C}\right)f && +r_2-r_3 &&=0\\
&-y_5 f && +3r_3 &&=0
\end{aligned} \quad (37)$$

and

$$\mathcal{P}=\left(P_P\left(\frac{y_4-C(1-y_5)}{1-C}\right)+P_D(1-y_4-y_5)\times\left(\frac{1-R}{1-C}\right)+P_G y_5-P_U\right)f-P_A f_1-P_B f_2-P_F \quad (38)$$

In this way $\mathcal{P}$ was calculated for the set of cost functions given by Williams, a mean annual cooling water temperature of 0.8125 (dimensionless units) and a whole range of values for f_2 and T. This was repeated for $f_1=0.1292$, 0.1512, and 0.1722, corresponding to the average flow rate of f_1 and its extreme variations. The results are shown in Fig. 3, and the contours of constant profit $\mathcal{P}$ in Fig. 4. When the general character of the profit surfaces was established, "hill-climbing" techniques were used to find the optimum settings of f_2 and T, and therefore of f_2 and f_w. These are shown plotted in Fig. 5 for a cooling water temperature of 0.8125. Similarly, the optimum settings were determined at a cooling water temperature of 0.828125 and an inlet stream temperature of 0.84375 (this corresponds to the highest average temperature); the results are included in Fig. 5.

Examination of the response surfaces shows them to be smooth unimodal functions departing but little from a quadratic surface. Hence steady-state maximization was especially easy. Maximum achievable $\mathcal{P}$ for various values of f_1 vary relatively little, and variations in the optimum settings of f_2 and T are also small. Yet the optimum settings for one value of f_1 yield surprisingly poor values of $\mathcal{P}$ for a different value of f_1. Thus, at $f_1=0.1512$, choosing $f_2=0.3500$ and $T=1.02$ yields $\mathcal{P}=0.383$. At $f_1=0.1292$, $f_2=0.3295$ and $T=1.01$ gives a $\mathcal{P}$ of 0.395. Yet, at $f_1=0.1292$, using $f_2=0.3500$ and $T=1.01$, $\mathcal{P}$ is 0.331, i.e., a 16% reduction in profit! This can be taken as an indication of the necessity of precise process control. At the same time, $\mathcal{P}$ seems to be increasing as f_1 decreases, indicating the desirability of a larger reactor, i.e., longer residence time. It must be borne in mind that these conclusions are subject to the assumptions made in deriving the model of the process. For the actual plant, such conclusions

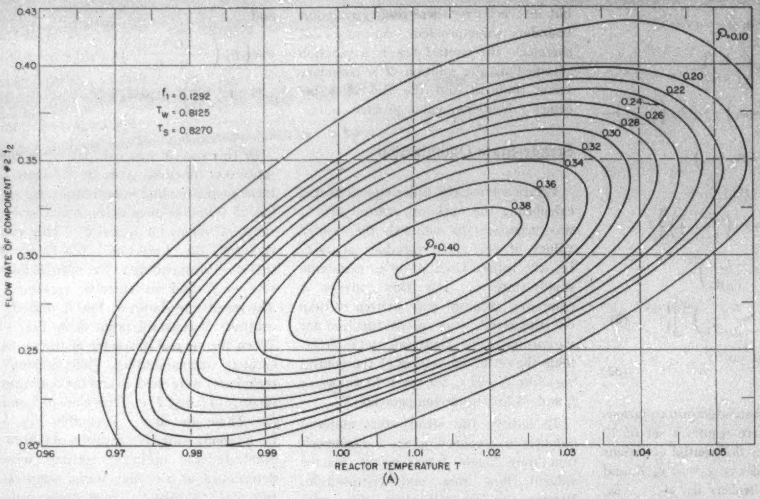


Fig. 4. Steady-state profit function contours A—At f_1= 0.1292. (For B and C see facing page)

may not be justified; they are presented here only as an indication of what can be determined in the initial stages of the optimization study. It must also be noticed that the optimum reactor conditions are not those specified by Williams, i.e., the profit functions do not agree with Figs. 20, 22, and 23 of his paper. One reason for this is that the azeotrope in the column was treated differently; another is that (in so far as is known to the authors) Williams' curves were obtained by varying one variable at a time, in particular by varying f_2 while keeping f_L fixed. In this problem f_L was tied to f_1 and f_2 by equations 33 and 36; hence the difference. Of course, for the conditions suggested by Williams, $f_1=0.1512$, $f_2=0.3475$, $T=1.00$, Fig. 4(B) indicates a $\mathcal{P}$ of 0.308 which is the same as shown by Williams. And yet the maximum $\mathcal{P}$ is found to be 0.383, an improvement of 24% or a monetary gain of $208,500 per year.

Calculation of the Control Law

The optimum control law for the reactor is given as a solution of the 2-point boundary value problem associated with the Euler equations; see equations 19 through 30. The boundary conditions are the values of the initial conditions of the system at $t=t_0$ and the optimum steady-state values of the variables at $t=t_e$. Therefore the controller has to choose a set of values, $\omega_1(t_0)$, $\omega_2(t_0)$, ... $\omega_6(t_0)$, which, for the existing state of the system $y_1(t_0)$, $y_2(t_0)$, ... $y_5(t_0)$, $T(t_0)$, will yield a solution of the Euler equation which satisfies the boundary conditions at $t=t_e$. Then, from the values of $\omega_i(t_0)$, $i=1,2$, ... 6, the controller can calculate the control variables $f_2(t_0)$ and $f_w(t_0)$. Unfortunately there is no direct method of choosing $\omega_i(t_0)$, $i=1, 2, \ldots 6$, since the equations are nonlinear and recourse to numerical trial-and-error solutions is necessary.

One possible method of solution makes use of the familiar hill-climbing techniques. The controller guesses a set of values of $\omega_i(t_0)$, $i=1, 2, \ldots 6$, and solves the Euler equations from $t=t_0$ to $t=t_e$. In general, the solution at $t=t_e$ will not satisfy the desired boundary conditions, so some criterion must be established to indicate how close to the desired point the solution passes. Since the primary objective of the solution is the maximization of $\mathcal{P}$, it is not reasonable to use the profit accumulated between t_0 and t_e as the indicator. Thus one attempts to find $\omega_i(t_0)$, $i=1, 2, \ldots 6$, to maximize

$$\mathcal{P}_a=\int_{t_0}^{t_e}(\mathcal{S}-\mathcal{C})\,dt$$

where the time functions y_1, y_2, ... y_5, T are the solution of the Euler equations with initial conditions $y_1(t_0)$, $y_2(t_0)$, ... $y_5(t_0)$, $T(t_0)$, $\omega_1(t_0)$, $\omega_2(t_0)$, ... $\omega_6(t_0)$. For sufficiently large t_e the values of $\omega_1(t_0)$, $\omega_2(t_0)$, ... $\omega_6(t_0)$ which maximize $\mathcal{P}_a$ approach those for which the solution of the Euler equations satisfies the desired end-point boundary conditions.

In this way the problem of finding functions $f_2(t)$, $f_w(t)$, $t_0 \leq t \leq t_e$, which maximize a functional, is reduced to a problem of finding variables to maximize a function, and this can be solved by any one of many hill-climbing techniques. Both the hill climbing and the solution of the Euler equations can be done "on line" by any of the modern computers.[2,3]

The above method of solution is a time-consuming process and would not normally be employed. Instead, one would obtain many solutions of the boundary-value problem and tabulate the values of $\omega_i(t_0)$, $i=1, 2, \ldots 6$, and therefore $f_2(t_0)$ and $f_w(t_0)$, as functions of $y_1(t_0)$, $y_2(t_0)$, ... $y_5(t_0)$, $T(t_0)$. To do this it has been found most efficient to generate solutions of the Euler equations from $t=t_e$ backward in time toward t_0. If $y_1(t_e)$, $y_2(t_e)$, ... $y_5(t_e)$, $T(t_e)$ are the set desired terminal conditions, then, for any set of $\omega_1(t_e)$, $\omega_2(t_e)$, ... $\omega_6(t_e)$, a solution will be generated which yields some $y_1(t_0)$, $y_2(t_0)$, ... $y_5(t_0)$, $T(t_0)$, $\omega_1(t_0)$, $\omega_2(t_0)$, ... $\omega_6(t_0)$, and the corresponding $f_2(t_0)$, $f_w(t_0)$. For this set of $y_1(t_0)$, $y_2(t_0)$, ... $y_5(t_0)$, $T(t_0)$ the corresponding $f_2(t_0)$ and $f_w(t_0)$ are the optimum settings of the control variables and so a point has been tabulated for the optimum control law functions

$$f_{2\,\mathrm{opt}}(t_0)=F_2[y_1(t_0), y_2(t_0), \ldots y_5(t_0), T(t_0)] \tag{39}$$

$$f_{w\,\mathrm{opt}}(t_0)=F_w[y_1(t_0), y_2(t_0), \ldots y_5(t_0), T(t_0)] \tag{40}$$

Thus, every time the Euler equations are solved, a point of F_2 and F_w is gen-

Fig. 4. B—At $f_1=0.1512$

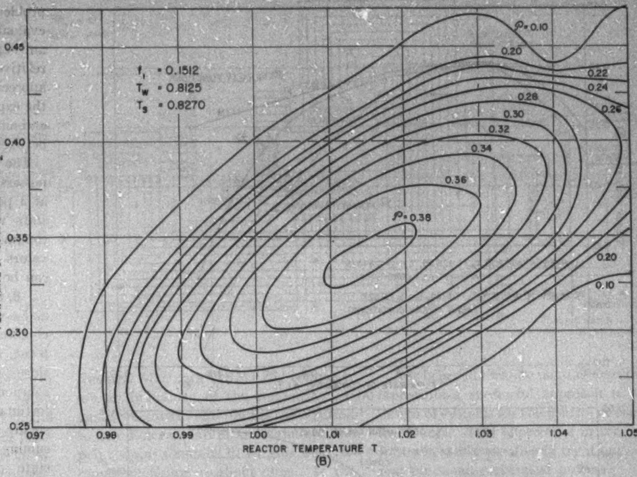
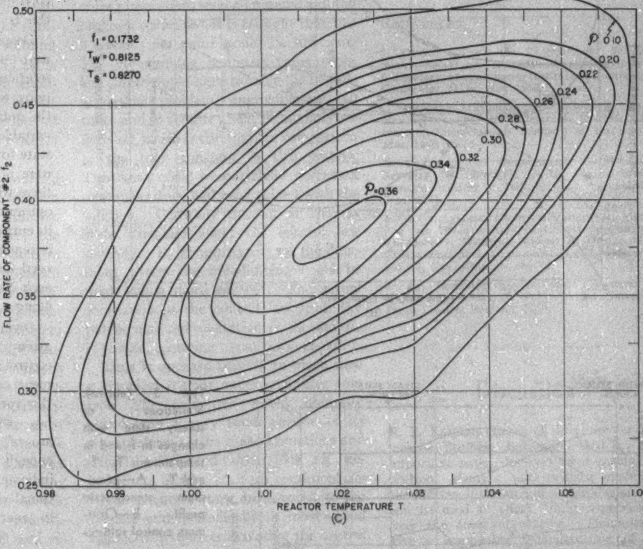


Fig. 4. C—At $f_1=0.1732$



erated and in this way the complete tabulation can be obtained. Needless to say, the number of entries may be very large, but in such a case one would obtain polynomial approximations to F_2 and F_w, thereby characterizing them by a relatively small number of parameters.

Once the functions have been obtained, the controller measures $y_1(t_0)$, $y_2(t_0)$, ... $y_5(t_0)$, $T(t_0)$, and from these generates $f_2(t_0)$ and $f_w(t_0)$ according to F_2 and F_w. For use by a digital controller F_2 and F_w may be represented in terms of the coefficients of the describing polynomials; for use by an analog controller these may be synthesized by analog function generators.

Tabulation of the control function, though simple in principle, is nevertheless difficult to carry out in practice. The Euler equations are always inherently unstable whether solved forward or backward in time.[1-3] Small changes in the initial conditions at $t=t_s$ cause large changes in the values of the solution at $t=t_0$, so it is not easy to spot points in the region of the domain of F_2 and F_w that are of interest. This is further aggravated by the presence of constraint functions F with sharp cutoff characteristics. In the problem being presented this leads to the evaluation of $f_2+f_1+f_L$ as the 15th root of an expression involving differences of relatively large numbers. If $f_2+f_1+f_L$ is appreciably less than 1, then the value of the expression must be very small indeed and must be obtained as a small difference between large numbers.

The Euler equations are known to possess a saddle-point type of singularity at a point in the phase space for which state variables yield a steady-state optimum.[8] If the steady-state optimum values are known, Lagrange multipliers can be chosen to make $\omega_i(t_e)=0$, $i=1,2, \ldots 6$, thereby defining the singular point, and so all the conditions at $t=t_e$. By choosing small perturbations from this point, solutions which diverge relatively slowly are obtained, thereby keeping the solutions in the domain of interest. Unfortunately, in the current problem the steady-state optimum was chosen assuming f_L tied to f_1 and f_2. The steady-state optimum of the Euler equations corresponds to an independent f_L; hence this trick cannot be used.

Despite these difficulties, solution points were obtained and the corresponding solutions are shown in Fig. 6. The solutions, when plotted forward in time, show the optimum path along which the state variables would move from the current state to the steady-state optimum if there were no disturbances. Needless to say, there will always be disturbances and the optimum path will never be followed in its entirety until the steady-state optimum is reached (otherwise there would be no need for dynamic control), but following each disturbance the system is redirected along a new optimum path.

Owing to lack of computing time, no attempt was made to compile the entire control law table. Therefore no statement can be made about how severely the various variables affect the control law—which play a minor role and which are of predominant importance. Only enough solution points were obtained to indicate the feasibility of this computational scheme and the control approach in general.

Synthesis of Control System

The design procedure outlined above leads to the formulation and evaluation of the optimum control law for the system. The designer is allowed the freedom of choosing any system configuration to realize this law. Presented below is one such configuration for the control of the stirred-tank reactor, with a few comments on the problems of realization.

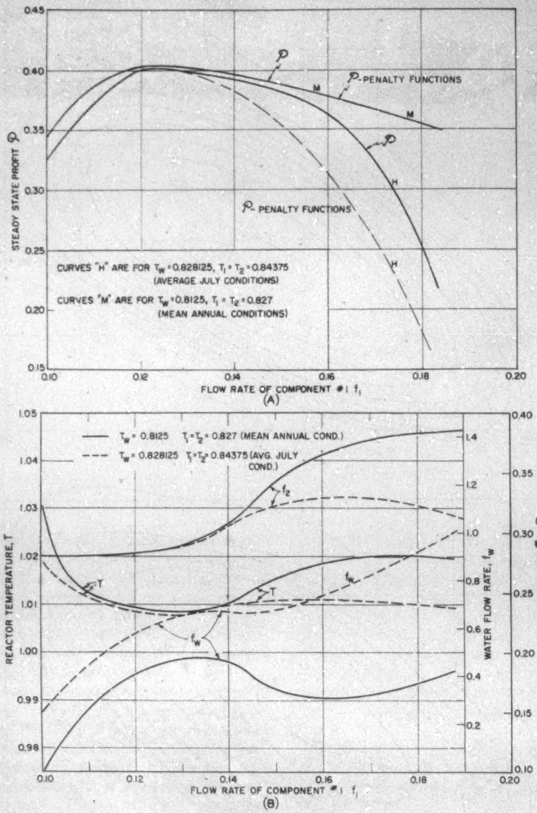


Fig. 5 (above). Variations of steady state with changes in f_1 and in temperatures T_w, T_1, and T_2. A—Maximum steady-state profit. B—Optimum control settings

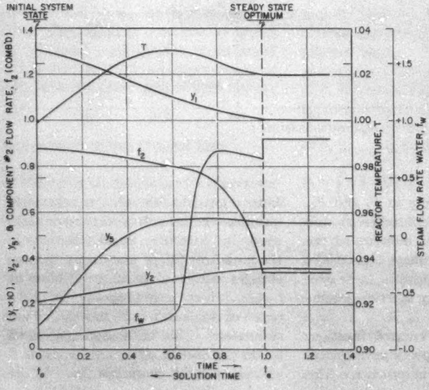


Fig. 6 (left). Sample solutions of Euler equations

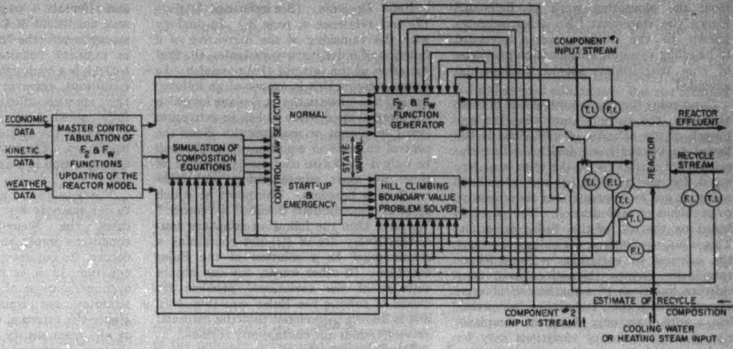

ECONOMIC DATA
KINETIC DATA
WEATHER DATA
MASTER CONTROL TABULATION OF F_2 & F_W FUNCTIONS UPDATING OF THE REACTOR MODEL
SIMULATION OF COMPOSITION EQUATIONS
CONTROL LAW SELECTOR
NORMAL
START-UP & EMERGENCY
STATE VARIABL
F_2 & F_W FUNCTION GENERATOR
HILL CLIMBING BOUNDARY VALUE PROBLEM SOLVER
COMPONENT #1 INPUT STREAM
T.I.
F.I.
REACTOR
REACTOR EFFLUENT
RECYCLE STREAM
ESTIMATE OF RECYCLE COMPOSITION
COMPONENT #2 INPUT STREAM
COOLING WATER OR HEATING STEAM INPUT

Fig. 7. Information flow in reactor control system

First, the control law is a function of the state variables $y_1(t_0)$, $y_2(t_0)$, ... $y_6(t_0)$. Ideally, composition measurements should be available to the controller but in practice this is very difficult. However, knowing reaction kinetics, reactor temperature, inlet and outlet flow rates, and composition, one may simulate the composition equations 4 to keep a running account of reactor composition. This is possible because the composition equations are stable and solution errors do not accumulate. Hence the simulation of the reactor equations must be an integral part of the controller.

The controller itself is the F_2 and F_w function generator; however, the domain of the function tabulation is limited. During start-up and emergencies, the state variables may be outside this domain and then the solution must be obtained from the boundary-value problem itself. Thus, in parallel with the F_2 and F_w function generator, a boundary-value problem solver must be incorporated. Since solution for a single set of state variables is required, the hill-climbing method of solution described in the section "Calculation of Control Law" would be most appropriate. The system would then be under the over-all control of a master monitor, which, on the basis of incoming economic data, results of further kinetic studies, and possibly weather information, updates the reactor model and recalculates the control law functions. Such an over-all control scheme and the information flow involved are shown in the diagram of Fig. 7.

Whether the over-all system is to be realized in terms of digital or analog components or in terms of a hybrid combination is an open question. Each method has advantages and disadvantages. The solutions of the control law as carried out on the International Business Machines (IBM) *704* computer are an indication of the computing capacity required. To obtain the solution to the Euler equation and tabulate one point, approximately 10 seconds were required including read-out time. If the output operations were eliminated and the program optimized, it is thought that 1 to 2 seconds per point would be sufficient. The hill-climbing boundary-value problem solver would have to solve the boundary-value problem approximately 40 times and so require about 80 seconds per decision, which is still small in comparison to the time constant of this system. This figure could be materially decreased once experience had been gained with the solution. Certainly not all of the six state variables affect the control law severely. By concentrating on the dominant effects the hill-climbing could be made much more efficient. The actual generation of the control law from the polynomial representation of F_2 and F_w is simple, requiring about 0.01 second.

Thus it appears feasible to place such a chemical reactor under computer control. The computer would normally spend only a very small portion of its time performing the control functions and would be free to do other jobs, e.g., accounting. Under start-up conditions its whole attention must be devoted to the process and, periodically, it must spend appreciable time recalculating the control laws.

It should be emphasized that the problem considered is 6-dimensional. No other schemes known to the authors were successful in handling nonlinear dynamic optimization problems of this many dimensions. Probably some of the variables are not important in affecting the control law; if this is true, tabulation of the control law function will make it evident.

It is believed that the method described represents a successful approach to the design of systems for the control of chemical plants. It is hoped that in the near future it will be possible to try it out on a real functioning chemical process.

References

1. Optimum Nonlinear Controllers, W. Kipiniak. *Report no. 7793-R2*, Electronic Systems Laboratory, Massachusetts Institute of Technology, Cambridge, Mass., Nov. 1958.

2. Progress Memorandum, W. Kipiniak. *Memorandum no. 7967-M1*, Electronic Systems Laboratory, Massachusetts Institute of Technology, Mar. 1960.

3. Control and Optimization of Dynamic Systems, W. Kipiniak. *Sc.D. Thesis*, Massachusetts Institute of Technology (to be published Jan. 1961).

4. A Generalized Chemical Processing Model for the Investigation of Computer Control, Theodore J. Williams, Robert E. Otto. *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 79, Nov. 1960, pp. 458-73.

5. An Introduction to the Calculus of Variations (book), C. Fox. Oxford University Press, London, England, 1950.

◆

Discussion

R. E. Kalman (Research Institute for Advanced Studies, Baltimore, Md.): This important paper deserves most careful attention. Using a quite realistic model for the process, it shows how a digital computer can be used to effect "tight supervision" over the chemical reactor. The authors give a convincing demonstration of the feasibility of computer control, a major step forward in research on process control which should impress engineers who like to "evaluate" research in terms of economic criteria.

Essentially the same problem, based on a simpler mathematical model, has been treated before by the discusser;[1] another application of the same technique has been described by R. E. Boydston, and was presented at the same session of the Joint Automatic Control Conference as the paper under discussion. Much of this stems from the pioneering work of Bellman.[3] Hence the statement "No other schemes known to the authors were successful in handling dynamic optimization problems of this magnitude" should perhaps be amended.

Concerning the mathematical side of their approach, the authors state that the approach of the calculus of variations is believed to be the most fruitful. But the calculus of variations, as the term is understood by mathematicians and physicists today, is not *an* approach but *the* approach to problems of optimization. The paper is based on the use of the Euler equation. This is but one (in fact, the oldest) aspect of the calculus of variations. The authors specifically exempt themselves from considering questions of a mathematical type associated with their technique. Nevertheless, a note of warning about the complexities of the calculus of variations may be useful at this stage, since much further work can be expected along this or similar lines.

The expositions of the calculus of variations usually found in engineering texts are badly lacking in rigor. In problems of classical theoretical physics, the rigorous details can usually be provided. This is not as easy in new types of problems such as occur in this paper. The following discussion is intended to point out that even in innocent-looking cases the rigorous application of the calculus of variations is far from being as simple as might appear.

There are only two modern, rigorous expositions in English to the best of my knowledge: Bliss[3] and the first two chapters of Morse.[4] These are not intended for the beginner but they contain a streamlined account of classical results.

Intuitively, the calculus of variations can be compared with finding the minima and maxima of a function $f(x_1, \ldots, x_n)$ of a finite number of variables. It is necessary that $\partial f/\partial x_i=0$ $(i=1, \ldots, n)$; the additional condition that the Hessian matrix $\partial^2 f/\partial x_i \partial x_j$ is negative (or positive) definite is then sufficient for a local maximum (or minimum). (The absolute maximum or minimum cannot be found by such simple means.) There are corresponding conditions in the calculus of variations which we shall now proceed to sketch.

Consider the problem of minimizing or maximizing the integral

$$V=\int_{t_0}^{t_1} L(x, \dot{x}, t)dt=\frac{1}{2}\int_{t_0}^{t_1} [x^2(t)\pm \dot{x}^2(t)]dt \quad (41)$$

subject to the constraint

$$\dot{x}=u(t) \quad (42)$$

Equation 42 represents a dynamical system: x is the state and u is the control variable; equation 41 defines the performance index. In general, though not always, one prescribes the initial and end points $x(t_0)$ and $x(t_1)$ of the desired motion of the system.

With the plus sign in equation 41 we are dealing with a conventional optimization problem. With the minus sign, equation 41 is intended to be a highly simplified version of equation 1: x^2 is the sales price of the output and u^2 is the price of the plant input.

Euler Equation. (See reference 3, page 12, and reference 4, page 3.) In analogy with the vanishing of the derivative of a function of finitely many variables, the first variation of the integral 41 is to vanish. In rigorous terms this is expressed as follows: Let $\xi(t)$ be a curve in the (x, t) plane for which the integral in equation 41 has an extremum (local minimum or maximum). Since the control variable in equation 42 might well be only a piecewise continuous function of time, we must admit curves with discontinuous tangents. The point at which $d\xi/dt$ is discontinuous is called a "corner." We have then the following necessary condition: Every arc of $\xi(t)$ not containing a corner must be a solution of the Euler equation. In other words, one can obtain extrema of the variational problem by means of solving the Euler equation only if it is known beforehand that the extremal curve $\xi(t)$ does not contain any corners.

The Euler equation is

$$\frac{d}{dt}(\partial L/\partial \dot{x})-\partial L/\partial x=0$$

or, for equation 41,

$$\pm d^2x/dt^2-x=0 \quad (43)$$

Weierstrass-Erdmann Corner Conditions. (See reference 3, page 12 and reference 4, page 3.) At a corner t_0 the extremal curve must satisfy the following condition:

$$\lim_{\tau\to 0} \partial L(\xi(t_0+\tau), \dot{\xi}(t_0+\tau), t_0+\tau)/\partial \dot{x}$$

$$= \lim_{\tau\to 0} \partial L(\xi(t_0-\tau), \dot{\xi}(t_0-\tau), t_0-\tau)/\partial \dot{x}$$

Legendre Condition. (See reference 3, page 23, and reference 4, page 6.) In general, the "loss function" L in the integral in equation 41 is a function of the vector variables $\mathbf{x}$ and $\dot{\mathbf{x}}$. Denoting the components of $\dot{\mathbf{x}}$ by $\dot{x}_1, \ldots \dot{x}_n$, can show that the integral will not have even a local maximum (or minimum) unless the matrix $\partial^2 L/\partial \dot{x}_i \partial \dot{x}_j$ is nonpositive (or nonnegative) definite. It should be emphasized that the Legendre condition must be considered in addition to solving the Euler equation.

Jacobi Condition. (See reference 3, page 26, and reference 4, page 8). It may happen that the second variation along any presumed extremal curve $\xi(t)$ (i.e., a solution of the Euler equation) vanishes. This can be shown to be equivalent to the following situation:

The Euler equation of the variational problem associated with the second variation has a solution which is zero at $t=t_0$ and $t=t_2>t_0$ but is not identically zero anywhere in the interval (t_0, t_2).

The point t_2 is called the conjugate point of t_0. The necessary condition of Jacobi states:

No curve $\xi(t)$ without corners, which is the solution of the Euler equation, can provide a (local) minimum or maximum if t_0 has a conjugate point in the interval (t_0, t_1).

As a matter of fact, one can show that if there is a conjugate point on $\xi(t)$, by introducing corners one can always modify $\xi(t)$ in such a way that the integral in equation 41 will assume a smaller or larger value.

It can be shown (see reference 3, page 43, and reference 4, page 16) that if the necessary conditions of Legendre and Jacobi are strengthened (the Hessian must be positive or negative definite, and no t satisfying $t_0<t\leq t_1$ is a conjugate point of t_0), then these conditions, together with the Euler equation, are also sufficient. It should be carefully noted that none of these results guarantee the existence of a minimum or maximum. The investigation of the existence question is a deep mathematical problem.

Now let us examine how these facts relate to our example and to equation 15. In all cases, the Weierstrass-Erdman corner conditions imply that there cannot be a corner. In our example this is obvious. In equation 15 it is brought about by the authors' method of "simulating" the constraints; see equations 11 through 13. Hence the extremal curves $\xi(t)$ (if they exist at all) must satisfy the Euler equation in the entire interval (t_0, t_1). The remaining part of the discussion will be carried out case by case.

Plus Sign in Equation 41. The strict Legendre condition is satisfied and shows that we can have only minima. The maximization problem has no solution (i.e., the maximum is infinite). The Euler equation can be shown have a unique solution connecting the arbitrary points $x(t_0)$ to $x(t_1)$ in any time interval (t_0, t_1). The proof of this is not trivial and leads to the concept of "controllability" introduced by the discusser.[5] This being the case, it is easy to prove also that there are no conjugate points. In fact, since L is quadratic, the Euler equation for the second variation is identical with equation 43; it follows that the only solution of this equation which links $x(t_0)=0$ with $x(t_1)=0$ (t_0, t_1 arbitrary) is the curve $\xi(t)\equiv 0$. More directly, one sees this by noting that, since L is positive definite, the second variation can never vanish. This is a case frequently encountered in classical physics. Since the three necessary conditions are also sufficient, it follows that the unique solution of the Euler equation is indeed the desired answer. Hence in this case one simply solves the Euler equation, forgets about all else, and comes out with a correct answer. The situation is different in the next case.

Minus Sign in Equation 41. The strict Legendre condition is satisfied and shows that we can have only maxima. The Euler equation can again be shown to have a unique solution corresponding to any pair of points $x(t_0,)$ $x(t_1)$, except for the pair 0,0, which can be connected by infinitely many solutions when $t_1-t_0=n\pi$ (n = positive integer). Since the Euler equation of second variation is again the same as equation 43, we see that t_0 has its conjugate points at $t=t_0\mp n\pi$. By the Jacobi condition, the unique solution of the Euler equation is a local maximum if $t_2<t_0+\pi$. But, if $t_1>t_0+\pi$, no curve $\xi(t)$ without corners can supply a maximum because of the presence of the conjugate point; on the other hand, curves with corners are ruled out for other reasons. Hence, if $t_1>t_0+\pi$, no maximum exists. One can show this directly as follows: Let $t_1=t_0+3\pi/2$. Connect $x(t_0)$ to any arbitrary point $y=x(t_0+3\pi/4)$ by means of the (unique) solution of the Euler equation. Similarly, connect $x(t_0+3\pi/4)$ to $x(t_1)$. Then the integral V in equation 41 will be a function of y. Take, e.g., $x(t_0)=1$ and $x(t_1)=0$. By an easy calculation, one sees that $V(y)\rightarrow\infty$ with $y\rightarrow\infty$, so that no maximum exists.

Equation 15. Even the verification of the Legendre condition is not a trivial matter here, but is probably true. On the other hand, the integrand in equation 15 is apparently not negative definite, so that conjugate points are not obviously ruled out. It may well be that the conjugate points have failed to show up only because of the relatively short time interval involved in the optimization; see Fig. 7. Unfortunately, it is difficult to give criteria as to how short this interval must be. Another question, by no means obvious, is: How do we know that any initial state in which the reactor might find itself can be transferred to the desired operating point in a certain fixed length of time? It is well known in the theory of saturating control systems that this is not always possible.

The difficulties sketched here have long been known and have led mathematicians to search for other approaches. The chief competitor or the Euler equation in this regard is the Hamilton-Jacobi theory, in particular, the partial differential equation bearing the same name. The existence of a solution of the Hamilton-Jacobi partial differential equation (with slight additions) guarantees the existence of a solution for the variational problem; as we have just seen, the same is by no means true of the Euler equation. In recent years, the Hamilton-Jacobi theory received splendid reaffirmation by the work of Bellman,[2] whose "dynamic programing" or "functional-equation" technique is based on the same general idea.[6]

The Hamilton-Jacobi equation for the example can be derived from the well-known "maximum" principle of Pontryagin.[6] Let p be the conjugate variable of x. We can eliminate equation 42 and write

$$L(x, u, t)=(x^2-u^2)/2=(x^2-\dot{x}^2)/2$$

(This elimination can be also carried out in the general case, obviating the authors' technique of considering the dynamical equations as constraints.) By virtue of the maximum principle, we select u in such a way as to minimize or maximize the Hamiltonian function

$$H(x, p, u, t)=L(x, u, t)+p\dot{x}=(x^2\pm u^2+2pu)/2$$

which yields

$$u_{\text{opt}}=k(x, p)=\mp p \tag{44}$$

The function k represents the "optimal control law." It specifies the mathematical form of the control system if one succeeds in expressing p as a function of x. (Incidentally, this derivation, with minor modifications, succeeds also when the control variables are constrained in amplitude,[6] which shows that the authors' belief that "inequality constraints cannot be readily treated by the techniques of the calculus of variations" appears to be incorrect.)

We now define the function

$$H^*(x, p, t)=H(x, p, u_{\text{opt}}, t)=(x^2\mp p^2)/2$$

The Euler equation 43 is then seen to be equivalent to the canonical (Hamiltonian) equations

$$\left.\begin{aligned}\dot{x}&=\partial H^*/\partial p=\mp p\\ \dot{p}&=-\partial H^*/\partial x=x\end{aligned}\right\} \tag{45}$$

We may consider V as a function of the initial point ξ_0, t_0 of the extremal curve. Since the initial point and time are arbitrary, we may drop the subscript zero, writing $V(\xi, t)$. Identifying p with $\partial V(\xi, t)/\partial\xi$, we are led to the Hamilton-Jacobi partial differential equation

$$\partial V/\partial t+H^*(x, \partial V/\partial\xi)=0$$

which in the example is given by

$$\partial V/\partial t+x^2/2\mp(\partial V/\partial\xi)^2/2=0 \tag{46}$$

Let

$$V(x, t)=\nu(t)\xi^2/2.$$

Substitution shows that this function satisfies equation 46 if the scalar ν is a solution of the ordinary differential equation of the Riccati type

$$-d\nu/dt=(1\mp\nu^2)/2 \tag{47}$$

The initial conditions of equation 47 are usually specified by letting $V(\xi, t_1)=0$. (This corresponds to a "free" end point of the motion at $t=t_1$; the case of "fixed" end point can be treated by similar methods.)

In case of equation 47, the initial condition is then $\nu(t_1)=0$. If the minus sign holds in equation 47 (i.e., the plus sign holds in equation 41), it can be readily shown that the unique solution of equation 47 starting at $\nu(t_1)=0$ exists for all $t<t_1$. This is not the case when the plus sign holds in equation 47 (i.e., when the minus sign holds in equation 41), because then the solution $\nu(t)$ will become infinite at some value of $t<t_1$. In fact, the solution of equation 47 is then given by

$$\nu(t)=\tan^{-1}(t_1-t)/2$$

and this solution becomes infinite when $t=t_1-\pi$, i.e., at the first conjugate point of t_1 on the left. We see therefore that the Hamilton-Jacobi technique allows one to conclude in this case the nonexistence of a solution for $t_1-t_0>\pi$, without any detailed knowledge of the classical necessary conditions.

One further point of both theoretical and practical importance: In the classical theory, one considers minimization of the integral in equation 41, with a loss function of type $L(x,\dot{x},t)$, and subject to a dynamic constraint, such as equation 42. But it is much simpler to use a loss function of type $L(x,u,t)$, in which case the differential equations of the plant are no longer constraints but are automatically considered in the derivation of the Hamiltonian. The LaGrange multipliers ω_i of equation 16 are then recognized as the conjugate variables p_i; the "maximum principle" of Pontryagin[6] follows naturally and the whole theory is greatly simplified. Details may be found in the discusser's forthcoming paper in the *Proceedings* of the RAND Symposium on Mathematical Optimization.

We emphasize again that if a solution of the Hamilton-Jacobi equation satisfying the appropriate initial conditions has been found, the variational problem is solved. One may regard solution of the Euler equation as an attempt to construct solutions of the Hamilton-Jacobi equation. (This is obvious from the fact that the Euler equations (in the canonical form 5 given by Hamilton) are just the equations for the characteristics of the Hamilton-Jacobi partial differential equation.) But the solution of the Euler equation does not automatically imply that there is a solution for the Hamilton-Jacobi equation (except for sufficiently short intervals of time) because of the conjugate point trouble. Hence in regard to theoretical questions, the Hamiltonian point of view is intrinsically the superior of the two.

The solution of the Hamilton-Jacobi equation completely specifies p as a function of the state; therefore, by equation 44, it also specifies the optimal control law as a function of the state. This may be more, however, than is really desired. It may be "cheaper" to specify the optimal control only for the given state of the system, rather than for all states. Under such circumstances, repetitive, real-time solution of the Euler equation, as used by the authors, may have appreciably advantages, provided of course that the Euler equation indeed provides the solution for the variational problem. In other cases, it might be better to solve the Hamilton-Jacobi equation directly, in terms of a family of orthogonal functions. In any case, one should guard against too hasty generalizations concerning the "engineering" merits of the Euler point of view.

The foregoing may be summarized roughly as follows:

1. Solution of the Euler equation provides a necessary but not a sufficient condition.

2. The solution of the Hamilton-Jacobi equation (or, equivalently, dynamic programing) provides a condition which, with slight additional requirements, is also necessary. This point of view is used in reference 1 and also by Boydston.

3. The computational advantages of the Euler point of view depend on the problem at hand and the requirements of control. It is primarily preferable in cases where the control law is to be determined step by step as the need arises, rather than all at once.

4. A brute-force numerical attack on the problem is bound to run into difficulties unless the existence of the optimal system is established in advance. It is by no means obvious that an extremum of the integral exists for any time interval (see equation 41 with minus sign) or that any state can be transferred to the desired final state.

None of the difficulties raised is insurmountable from the engineering point of view, but neither is the problem as simple as might be surmised by a quick reading of the paper.

REFERENCES

1. THE ROLE OF DIGITAL COMPUTERS IN THE DYNAMIC OPTIMIZATION OF CHEMICAL REACTORS, **R. E. Kalman, R. W. Koepcke.** *Proceedings*, Western Joint Computer Conference, Institute of Radio Engineers, New York, N. Y., 1959, pp. 107–16.

2. DYNAMIC PROGRAMMING (book), **R. E. Bellman.** Princeton University Press, Princeton, N. J., 1957.

3. LECTURES ON THE CALCULUS OF VARIATIONS (book), **G. A. Bliss.** University of Chicago Press, Chicago, Ill., 1946.

4. THE CALCULUS OF VARIATIONS IN THE LARGE (book), **M. Morse.** American Mathematical Society, New York, N. Y., 1934.

5. ON THE GENERAL THEORY OF CONTROL SYSTEMS, R. E. Kalman. *Proceedings*, First International Congress on Automatic Control, Moscow, USSR, 1960.

6. L. S. PONTRYAGIN'S MAXIMUM PRINCIPLE IN THE THEORY OF OPTIMUM SYSTEMS, I-III, L. I. Rozonoër. *Automatika i Telemekhanika*, Moscow, USSR, Oct.-Dec. 1959.

L. A. Gould and **W. Kipiniak:** The authors appreciate Dr. Kalman's comments. There is no question that the method requires a rigorous foundation and that care is needed in applying the technique. This paper was presented to show how the method could be used on a realistic engineering problem, in a situation in which the questions of existence and uniqueness of an optimum can be resolved on physical grounds. At present, not all of the mathematical restrictions on the application of the method are completely known; investigation of these restrictions is now under way.

With regard to the example presented in the discussion in which the Euler equations did not define an optimum, we should point out that it does not represent a realistic control problem. In practice, the performance criterion represents a balance between the cost of tolerating an error and the cost of exerting control effort to correct it. That is, the integrand is either concave in both the state variables and their derivatives or convex in both. In Dr. Kalman's example in which a plus sign was used, the above criterion is satisfied and no difficulties result. The example with a minus sign corresponds to a criterion convex in the state variable but concave in its derivative and it is not surprising that the stationary point defined by the Euler equation is a point of inflection and not an optimum.

As far as is known to the authors (reference 1 of the discussion included), the dynamic optimization problems treated so far have been either of small dimensionality or based on a linear model and a quadratic performance criterion. We know of no other scheme successfully used for dynamic optimization of a nonlinear reactor model.

It must be emphasized that in contrast to Bellman's and Kalman's work, the optimization procedure is based on Euler differential equations of the calculus of variations rather than on the functional equations of dynamic programing. Although both approaches are, to a large extent, theoretically equivalent, computationally they present different problems. In our experience the variational approach has been found to possess some advantages.

The authors have always felt that all of the various optimization techniques have their place and that more applications of these methods are needed, to demonstrate their power both mathematically and in practical problems. This paper attempted to show one of these methods applied to a realistic engineering problem.

On the Response of Linear Systems to Periodic Excitations

W. R. LEPAGE
MEMBER AIEE

R. McFEE
MEMBER AIEE

THE PROBLEM of finding the periodic response of a linear system to a nonsinusoidal periodic excitation is of continuing interest.[1–6] The following constitutes a brief summary of the existing methods of solving such problems:

1. The Fourier series.
2. Infinite series summation of transients.
3. Solution of a differential equation subject to matching conditions at the beginning and end of a cycle.
4. Computation of the periodic response in closed form, for the duration of the first cycle, by subtracting the transient part of the response from the total response.
5. Use of the modified Z-transform to obtain the closed form solution for one cycle of response, in the limit, as t approaches infinity.

Methods 1 and 2 are of such classic nature as to need no comment, other than the well-known observation that they are seldom useful for computation. Method 3 is satisfactory only in simple cases, for which Laplace techniques are not needed. A variety of justifications for method 4 will be found in the literature[4–7] and method 5 has been expounded in a recent communication by Jury.[8] Method 5 is conceptually satisfying, to the degree that its procedure is closely identified with the definition of the periodic response as that part of the response which remains after the transient has died out. However, with respect to the amount of manipulation required to solve a specific problem, method 4 is probably not exceeded in simplicity by any of the other methods.

Attention will be devoted to presenting a new proof of the validity of the usual procedure for identifying the periodic and aperiodic parts of the response in method 4. All methods of justifying this procedure involve an interpretation of the inversion integral

$$y(t)=\frac{1}{2\pi j}\int_{Br} H(s)X(s)e^{st}ds$$

where $y(t)$ is the system response, $X(s)$ is the Laplace transform of the periodic driving function, $H(s)$ is the system function, and Br is the Bromwich contour. $X(s)$ has an infinite unbounded set of poles; and for a lumped system $H(s)$ is rational, having a finite number of poles. All procedures classified in method 4 recognize that residues due to poles of $H(s)$ constitute the transient response, and residues due to poles of $X(s)$ constitute the periodic response. However, a difficulty arises when the poles of $X(s)$ are used to identify the periodic response, because of the need to justify a term-by-term integration over an infinite set of poles. This problem has been attacked on the one hand by defining a special steady-state transform,[6] and also by using the conventional Laplace transform, while paying attention to the relatively difficult mathematical problem of justifying a term-by-term integration of an infinite series.[5]

It is the purpose of this paper to show that adequate theorems exist to demonstrate the validity of procedure 4 without defining a special transform, or without dealing with the infinite unbounded set of poles of $X(s)$ in the evaluation of the inversion integral.

The Basic Derivation

We are given a stable lumped linear system with a rational system function (ratio of Laplace transforms of output to input) $H(s)$. The system is initially without stored energy, and is excited at $t=0$ by a piecewise continuous periodic driving function $x(t)$, having period T. It is assumed that $x(t)$ and $H(s)$ meet conditions sufficient to ensure existence of the response function $y(t)$. Then $y(t)$ is given by

$$y(t)=\mathcal{L}^{-1}[H(s)X(s)] \tag{1}$$

The periodic input $x(t)$ has a transform[9]

$$\mathcal{L}[x(t)]=X(s)=\frac{X_0(s)}{1-e^{-sT}} \tag{2}$$

where by definition

Paper 60-1288, recommended by the AIEE Basic Sciences Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Chicago, Ill., October 9–14, 1960. Manuscript submitted July 12, 1960; made available for printing August 19, 1960.

W. R. LEPAGE and R. McFEE are with Syracuse University, Syracuse, N. Y.

Vol.
Pt. 1
196

$$X_0(s)=\mathfrak{L}[x_0(t)]=\int_0^T x(t)e^{-st}dt \qquad (3)$$

and where

$$\begin{aligned} x_0(t)&=x(t), \quad 0\leq t\leq T \\ &=0, \quad T<t \end{aligned} \qquad (4)$$

Since the integral in equation 3 has finite limits, and $x(t)$ is bounded, this integral is differentiable with respect to s for all finite s. Thus, $X_0(s)$ is an entire function, and $X(s)$ is singular only at the simple poles which arise from the zeros of $1-e^{-sT}$. These zeros occur on the imaginary axis at $\pm 2n\pi/T$, where $n=0, 1, 2, \ldots$. The transform of the output is

$$Y(s)=H(s)X(s)$$

or

$$Y(s)=\frac{H(s)X_0(s)}{1-e^{-sT}} \qquad (5)$$

$H(s)$ is a rational function, which is usually regular at infinity. However, there are certain degenerate cases for which $H(s)$ can have a pole at infinity. This possibility will be treated later; for the present, assume that $H(s)$ is regular at infinity.

The function $Y(s)$ has simple poles at the zeros of $1-e^{-sT}$ and is also singular at poles of $H(s)$, which are not necessarily simple. To distinguish between these two sets of poles, we now define $A(s)$ as the sum of the principal parts of $Y(s)$ due to the poles $s_1, s_2, \ldots s_N$ of $H(s)$. Then the function

$$P(s)=Y(s)-A(s) \qquad (6)$$

will have poles only at the zeros of $1-e^{-sT}$. These poles can be placed in evidence by writing

$$P(s)=\frac{P_0(s)}{1-e^{-sT}} \qquad (7)$$

where $P_0(s)$ is defined by equations 6 and 7, and may be written

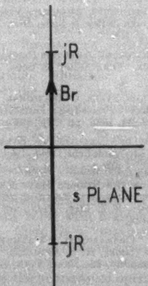


Fig. 1. Definition of the Br contour

$$P_0(s)=H(s)X_0(s)-A(s)+e^{-sT}A(s) \qquad (8)$$

Since $A(s)$ is a sum of a finite number of principal parts such as

$$\frac{b_m}{(s-s_1)^m}+\frac{b_{m-1}}{(s-s_1)^{m-1}}+\ldots+\frac{b_1}{s-s_1}$$

it follows that $A(s)$ has an inverse

$$a(t)=\mathfrak{L}^{-1}[A(s)] \qquad (9)$$

which is aperiodic if each pole has a nonzero real part. Since the physical conditions of the problem have been stipulated to ensure existence of $y(t)=\mathfrak{L}^{-1}[Y(s)]$, it follows that

$$p(t)=\mathfrak{L}^{-1}[P(s)] \qquad (10)$$

exists.

We suspect that $p(t)$ is periodic, and offer the following conjectural arguments in support of this statement. $P(s)$ is known to be singular only at the zeros of the denominator of equation 7, and so it is evident that $P_0(s)$ is an entire function. In view of the similarity with $X_0(s)$, which also is an entire function, and the similarity between equations 2 and 7, it is reasonable to expect $p(t)$ to be periodic.

To prove that $p(t)$ is periodic, we demonstrate first that, if $p_0(t)$ is zero for $t<0$ and $t>T$, then $p(t)=(pt-kT)$, where k is an integer; i.e., it is shown that the above conditions on $p_0(t)$ are sufficient to make $p(t)$ periodic. It is then shown that $p_0(t)$ meets the above conditions. Parenthetically, it is emphasized that we are now discussing sufficiency of a condition on $p_0(t)$ to ensure periodicity of $p(t)$. Necessity of the above condition on $p_0(t)$ has already been established in the discussion pertaining to equation 4.

On the basis of the above information, pending proof that $p(t)$ is periodic, the following may be written:

$$y(t)=a(t)+p(t) \qquad (11)$$

where $p(t)$ is periodic, and $a(t)$ is not periodic.

Proof that p(t) is Periodic

The following proves that $p(t)$ is periodic if $\mathfrak{L}^{-1}[P_0(s)]$ is zero when t is outside the interval from 0 to T:

To discuss $p(t)$, as given by equation 10, where $P(s)$ is given by equation 7, assume that $p_0(t)$ is zero when t is outside the interval from 0 to T. It is readily seen that the function

$$\sum_{k=0}^{\infty} p_0(t-kT)$$

is periodic, having the transform

$$\frac{P_0(s)}{1-e^{-sT}}$$

where $P_0(s)=\mathfrak{L}[p_0(t)]$. Thus, by virtue of uniqueness of transform pairs, it follows that

$$p(t)=\sum_{k=0}^{\infty} p_0(t-kT) \qquad (12)$$

is the inverse of $P(s)$, as given by equation 7, and is periodic if $\mathfrak{L}^{-1}[P_0(s)]$ exists and is zero outside the interval from 0 to T.

The following proves that $\mathfrak{L}^{-1}[P_0(s)]$ is zero when t is outside the interval from 0 to T:

The existence of $\mathfrak{L}^{-1}[P_0(s)]$ is assured because each term on the right of equation 8 has an inverse transform. Furtheremore, this inverse transform is given by the integral

$$p_0(t)=\frac{2}{2\pi j}\int_{Br} P_0(s)e^{sT}ds \qquad (13)$$

where Br is the path of integration obtained by allowing R to approach infinity in Fig. 1. In the general case of this inversion integral, path Br must be to the right of all singular points, but since $P_0(s)$ is an entire function, Br can coincide with the imaginary axis, as shown. Now consider the individual terms of equation 8.

To show that $p_0(t)$ is zero for $t<0$, it will be recalled that $x_0(t)=\mathfrak{L}^{-1}[X_0(s)]$ is zero for $t<0$; thus either from physical considerations, or the convolution integral

$$\int_0^t x_0(\tau)h(t-\tau)d_\tau$$

where $h(t)=\mathfrak{L}^{-1}[H(s)]$, it is evident that $\mathfrak{L}^{-1}[H(s)X_0(s)]$ is zero for $t<0$. Also, from standard Laplace transform theory, it is known that $\mathfrak{L}^{-1}[A(s)]$ and $\mathfrak{L}^{-1}[e^{-sT}A(s)]$ are both zero for $t<0$.

It is more difficult to prove $p_0(t)$ is zero

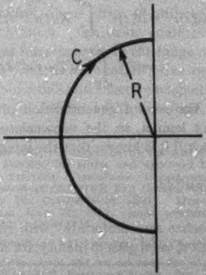


Fig. 2. A contour equivalent to Br when there are no left half-plane poles, and when conditions of Jordan's lemma apply to the left half-plane

for $t>T$. To accomplish this, we write the integral of equation 13 in the form

$$\int_{Br} P_0(s)e^{st}ds = \lim_{R\to\infty}\int_C P_0(s)e^{st}ds \tag{14}$$

where C is the semicircle of radius R shown in Fig. 2. Assuming that the limit on the right exists, this change is permitted because $P_0(s)$ has no finite-plane singular points. The nature of the singularity of $P_0(s)$ at infinity determines whether the limit exists. It can be shown that the right side of equation 14 is zero for $t>T$ by means of Jordan's lemma, which in one form states that

$$\lim_{R\to\infty}\int_C F(s)e^{su}ds=0$$

when $u>0$, if $F(s)$ approaches zero uniformly with respect to ang s, as $|s|$ goes to infinity, for $Re\ s \leqq 0$. Since we are interested in the response for $t>T$, it is convenient to write

$$\int_C P_0(s)e^{st}ds = \int_C P_0(s)e^{sT}e^{s(t-T)}ds$$

$$= \int_C P_0(s)e^{sT}e^{su}ds \tag{15}$$

Thus, in order to apply Jordan's lemma, we must show that

$$P_0(s)e^{sT} = H(s)X_0(s)e^{sT} - A(s)e^{sT} + A(s)$$

approaches zero as $|s|$ goes to infinity, uniformly in the region defined above. From the description of $A(s)$ as a sum of principal parts, it is evident that $A(s)$ goes to zero at least as fast as $1/s$, which goes to zero uniformly for $0 \leqq \text{angle } s \leqq 2\pi$. For the term $A(s)e^{sT}$, the exponential causes the region of uniform approach to zero to be reduced to $\pi/2 \leqq \theta \leqq 3\pi/2$ (or $Re\ s \leqq 0$).

To deal with the remaining term, observe that

$$H(s)X_0(s)e^{sT} = H(s)e^{sT}\int_0^T X_0(t)e^{-st}dt$$

$$= H(s)\int_0^T x_0(T-\tau)e^{s\tau}d_\tau \tag{16}$$

By the second generalization of Riemann's lemma, in the nomenclature of Doetsch,[10] it is known that the integral

$$\int_0^A f(x)e^{zx}dx$$

approaches zero uniformly* with respect to angle z, as $|z|$ goes to infinity, for $Re\ z \leqq 0$ provided that $f(x)$ is piecewise continuous. The factor $H(s)$ at most approaches a constant at infinity, and hence the last integral in equation 16 has the proper behavior at infinity. Thus, all terms of $P_0(s)e^{sT}$ approach zero in the proper manner for Jordan's lemma to be applicable; thus it follows that

$$\lim_{R\to\infty}\int_C P_0(s)e^{sT}e^{su}ds=0$$

when $u>0$, and hence, referring to equation 15

$$\lim_{R\to\infty}\int_C P_0(s)e^{st}ds=0$$

when $t>T$.

It has now been established that

$$\mathcal{L}^{-1}[P_0(s)]=0$$

when $t<0$, and when $t>T$. Furthermore, in the previous section it was shown that this condition, along with other conditions already established, is sufficient to make $p(t)$ periodic.

Case of H(s) with Pole at Infinity

As yet the proof is valid only for the case where $H(s)$ is regular at infinity. Now, suppose $H(s)$ has a kth order pole at infinity. Two cases are open to consideration:

1. The excitation is of such a form that the response includes impulse terms.
2. The excitation is smooth enough so that the response is free of impulse terms.

Case 1 can always be considered a limiting form of case 2, by using a symbolic notation where 1, s, s^2, etc., are interpreted respectively as symbolic transforms of the unit impulse, the unit doublet, etc. Therefore, we turn our attention to case 2.

It can be shown, either from physical or mathematical considerations, that if the response is to be free of impulses, $x_0(t)$ must have $k-1$ continuous derivatives, all of which will be zero at $t=0$. To simplify the analysis, assume the kth derivative as piecewise continuous. This being true, from the theorem for the transform of a derivative

$$X_0(s)=\frac{1}{s^N}X_k(s) \tag{17}$$

where

$$X_k(s)=\int_0^T \frac{d^k x_0(t)}{dt^k}e^{-st}dt \tag{18}$$

Note that $X_k(s)$ is an entire function, since the derivative is piecewise continuous. In place of equation 5 we can now write

$$Y(s)=\frac{H(s)}{s^k}\,\frac{X_k(s)}{1-e^{-sT}} \tag{19}$$

All arguments previously applied to $X_0(s)$ and $H(s)$ are now respectively applicable to $X_k(s)$ and $H(s)/s^k$. Since $X_k(s)/s^k = X_0(s)$, there is no difference between equations 5 and 19 in so far as the functions $A(s)$ and $P(s)$ derived from $Y(s)$ are concerned. Therefore, although equation 19 is conceptually necessary for carrying out the proof, in a numerical problem this case is handled by the procedure already described.

Summary

As pointed out at the start of the paper, the analysis presented leads to the well-established procedure for solving for a periodic response. The claim is made, however, that the subject becomes conceptually simpler with the realization of the following: In addition to the recognized fact that if $p(t)$ is periodic its transform has the form

$$\frac{P_0(s)}{1-e^{-sT}}$$

we can also say that if $P_0(s)$ meets certain conditions, the inverse transform of the above form exists and is periodic. This fact allows us to deal only with $P_0(s)$ and its inverse, thereby avoiding complexities arising from the infinite set of poles occurring at the zeros of $1-e^{-sT}$.

References

1. The Response of Electrical Networks to Non-Sinusoidal Periodic Waves, N. Marchand. *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 29, June 1941, pp. 330–33.

2. The Steady State Response of a Network to a Periodic Driving Force of Arbitrary Shape and Applications to Television Circuits, C. W. Carnahan. *Ibid.*, vol. 23, Nov. 1935, pp. 1393–1404.

3. The Response of a Network to an Arbitrary Periodic Driving Force, E. W. Hanlin. *Journal*, Franklin Institute, Philadelphia, Pa., vol. 233, 1942, pp. 257–70.

4. Linear Transient Analysis, Vol. 1 (book), E. Weber. John Wiley & Sons, Inc., New York, N. Y., 1954, p. 200.

5. Linear Network Analysis (book), S. Seshu, N. Balabanian. John Wiley & Sons, Inc., 1959, p. 168.

6. The Steady State Operational Calculus, D. L. Waidelich. *Proceedings*, Institute of Radio Engineers, vol. 34, 1946, pp. 78P–83P.

7. Response of Circuits to Steady State Pulses, D. L. Waidelich. *Ibid*., vol. 37, Dec. 1949, pp. 1396–1401.

8. A Note on the Steady State Response of Linear Time-Invariant Systems to General Periodic Input, E. I. Jury. *Ibid.*, vol. 48, no. 5, 1960, pp. 942–44.

9. Modern Operational Mathematics in Engineering (book), R. V. Churchill. McGraw-Hill Book Company, Inc., New York, N. Y., 1948, p. 54.

10. Handbuch der Laplace-Transformation Vol. 1 (book), G. Doetsch. Verlag Birkhauser, Basle, Switzerland, 1950, p. 170.

---

* Riemann's theorems have the physical interpretation that the frequency spectrum of any piecewise continuous function eventually becomes zero as frequency approaches infinity. However, this physical interpretation does not take cognizance of the uniform approach to zero, which is an important aspect of Riemann's lemma.

Discussion

E. I. Jury (University of California, Berkeley, Calif.): The authors are to be commended for their brief summary of methods and concise proof of the validity and ramifications of method 4. The discusser agrees that methods 4 and 5 show more promise in application than the others. In the following methods 4 and 5 are briefly compared, so that the reader may judge the difference between the various manipulations required to obtain the steady-state response. A brief summary of the steps involved in each of these methods is included.

In method 4, the process of obtaining the steady-state response denoted as y_{ss}, the following integrals which are equivalent to the inverse of equation 8 for $0 \leq t < T$, should be evaluated.[1]

$$y_{ss} = \frac{1}{2\pi j}\int_{C_1} H(s)X_0(s)e^{st}ds - \frac{1}{2\pi j}\int_{C_2} \frac{H(s)X_0(s)}{1-e^{-Ts}}e^{st}\,ds,\ 0 \leq t < T \quad (20)$$

The first integral is at the right of the poles of $H(s)X_0(s)$. The second integral encloses the poles of $H(s)$.

In evaluating the first and second integral the residues at the poles of $H(s)X_0(s)$ [for the various regions[1] of $X_0(s)$] and $H(s)$ are to be evaluated respectively. The steady-state output y_{ss} is obtained by subtracting the second integral (the transient response for the duration of the first cycle) from the first integral (the total response for the first cycle).

In method 5 with the use of the modified z-transform, the steady-state response is given.[3]

$$\begin{aligned} y_{ss} &= \lim_{z\to 1} \frac{z-1}{z}\left.\frac{Z_m[H(s)X_0(s)]}{(1-z^{-1})}\right|_{z=e^{Ts}} \\ &= \lim_{z\to 1} HX_0^*(z,m) \\ &= \lim_{z\to 1}\frac{1}{2\pi j}\int_{C_1} H(p)X_0(p)\left.\frac{e^{mpT}}{1-e^{-T(s-p)}}dp\right|_{z=e^{Ts}} \end{aligned}$$

$$0 \leq mT \leq T \quad (21)$$

The integral is at the right of the poles of $H(s)X_0(s)_{s=p}$.

To evaluate the integral of the modified z-transform $HX_0^*(z, m) = Z_m[H(s)X_0(s)]$, it should be observed that only the residues at the poles of $H(s)X_0(s)$ [for the various regions[2] of $X_0(s)$] are required.[2,3] Thus in this procedure the steady-state response is obtained in one step and by the evaluation of one integral.[3] The amount of manipulation required to solve for the steady-state using methods 4 and 5 can be easily compared by solving a specific problem. The limit $z = 1$ could also be inserted inside the integral in equation 21 which probably simplifies the manipulations required in obtaining the steady state.

Knowing the modified z-transform of $Y(s)$, one can readily obtain the total response (if desired) in a closed form using the following relation:[2,3]

$$y(n, m)T = \frac{1}{2\pi j}\int_{\Gamma} Y^*(z, m)z^n dz$$

The integral encloses the poles of $Y^*(z, m)$.

$t = (n+m)T$

n = positive integer

$0 \leq m \leq 1$

Furthermore, by expanding $zY^*(z, m)$ in powers of z^{-1}, the piecewise response for each period can also be obtained. As a final remark, utilizing the equivalence between methods 1 and 5, closed expressions of infinite Fourier series are obtainable.[3] Similarly, one could use the procedure of this paper for closing such series form.

In conclusion, the discusser deems it pertinent to mention that it is possible to combine the two integrals in equation 20 into a form equivalent to equation 21 (provided z is replaced by unity in the integrand of the latter). Thus method 4 reduces to method 5 or vice versa.

However, the availability in the literature[3] of extensive tables of the modified z-transform lends its application for obtaining the steady-state response conveniently.

REFERENCES

1. See reference 5 of the paper.
2. See reference 8 of the paper.
3. SAMPLED-DATA CONTROL SYSTEMS (book), E. I. Jury. John Wiley & Sons, Inc., New York, N. Y., 1958, chaps. 2 and 6.

W. R. LePage and **R. McFee**: The authors are glad to acknowledge Prof. Jury's comments, with which they find no significant disagreement. We wish to point out, however, that the discussion pertains only to our comment concerning the amount of work involved in solving a numerical problem. The main point of our paper is to show that the correctness of method 4 can be established without consideration of the difficult problem of evaluating the inversion integral when the integrand has an infinite set of poles.

Magnetic Memory Drum Design

E. B. CARNE
ASSOCIATE MEMBER AIEE

Synopsis: A magnetic memory drum is a convenient, relatively cheap storage device of moderate speed and of large capacity. It comprises a cylindrical rotor, whose surface is coated with a magnetic material, rotating within a shroud which supports salient heads. Proper design requires an appreciation of the effect on performance of such factors as rotor eccentricity, head-setting variations, head manufacturing tolerances, and storage surface variations. Equations are developed to describe the signal-to-noise ratio of a typical drum and the threshold (switching) voltage appropriate to the read amplifier for reliable operation. Material selection is approached on the basis of stress calculations and it is concluded that *2014-T6* aluminum is the best rotor material for general-purpose applications. It is shown that, for any rotor with a given degree of unbalance, there is a maximum stable operating speed. Problems associated with thermal mismatch of materials are discussed, and the necessity of operating large drums in a stable temperature environment is pointed out.

A MAGNETIC memory drum comprises a cylindrical magnetic surface rotating in close proximity to a number of magnetic heads. Information is stored on the surface by energizing a head with a suitable current pulse and producing a magnetized area on the rotating surface. Reading is achieved by noting the voltage produced at the output terminals when the magnetized area passes under a head. Because of the air gap between the head and the recording surface, the readback voltage is necessarily small; this presents a problem in terms of noise and spurious signal pickup when a large number of heads are involved. Nevertheless, magnetic memory drums have assumed an important place in data-processing systems, and are currently used where a convenient, potentially large, relatively cheap, permanent-storage device of moderate speed is required.

The storage surface may be fine iron-oxide particles held together by an organic binder or a very thin deposit of two or more metals, such as nickel and cobalt, plated simultaneously on the rotor cylinder. The former comprises micron-size powder in a resin matrix which, after appropriate curing, can be ground to a thickness of one-half to one mil (inches × 10^{-3}). The latter comprises a plated deposit, approximately 80% cobalt and 20% nickel, about 20 to 40 millionths of an inch thick. For oxide coatings, the coercive force may be in the range of 200 to 300 oersteds with an effective permeability of 2 to 3 and effective magnetic induction of 500 to 1,000 lines/cm² (per

---

Paper 60-1200, recommended by the AIEE Computing Devices Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Chicago, Ill., October 9-14, 1960. Manuscript submitted May 10, 1960; made available for printing August 3, 1960.

E. B. CARNE is with Melpar, Inc., Falls Church, Va.

The author would like to acknowledge the help rendered by A. S. Berson, of Melpar, Inc., in the discussion of electrical results and in providing the experimental values of head crosstalk voltages; also the assistance of J. A. Ward and F. A. Shultz, of the Remington Rand Division of the Sperry Rand Corporation, Utica, N. Y., in the formulation of the mechanical problem.

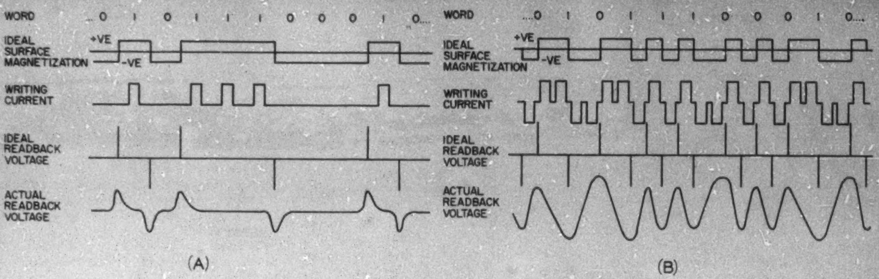


Fig. 1. Recording waveforms. A—NRZ. B—Phase-modulation

centimeter^2). For a typical Ni-Co (nickel-cobalt) deposit, the coercive force may be 200 to 800 oersteds, effective permeability 20, and magnetic induction 10,000 lines/cm^2.

In a typical system the air gap between the rotor and heads may be 0.001 inch. So that this dimension may be maintained with reasonable accuracy, the rotor diameter must be machined to a tolerance of ±0.0001 inch, and the mechanical stability of the bearings and housing must be perfect. A factor K_e will be introduced to represent the air-gap variation (rotor eccentricity) in such a way that

$$d_a' = d_{a0}(1 \pm K_e) \qquad 0 \leq K_e \leq 1 \tag{1}$$

where d_a' is the effective air gap between the storage surface and the head and d_{a0} is the nominal air gap. Expressed another way

$$D' = D_0 \pm d_{a0}K_e \tag{2}$$

where D' is the effective rotor diameter and D_0 is the nominal rotor diameter. Departures from the nominal air gap give rise to modulation of the maximum readback signal and variations in the length of the recorded pulse x_p.

A further important source of variation is the magnetic head used for writing and reading. K_g will be introduced to represent the variation in effective pole face gap, in such a way that

$$g' = g_0(1 \pm K_g) \qquad 0 \leq K_g \leq 1 \tag{3}$$

where g' is the nominal pole face gap. While it is usual to establish this dimension by means of nonmagnetic foil of closely controlled thickness, subsequent finishing operations can degrade the sharp corners and increase the effective gap width. This will affect the size of the magnetized area recorded on the storage surface.

Other discrepancies occur in the manufacture of the head, such as variation of the number of turns of wire in the energizing coil and variation of the parameters of the head magnetic circuit material. A factor K_m will be introduced to represent the effect of these manufacturing tolerances. There will also be a variation in the setting of the head relative to the storage surface. A factor K_a will be used to represent this kind of variation. Its value will depend on the style of head used, and on the way in which the heads are set. Finally, the magnetic playback will depend on the quality of the magnetic surface, which will depend very critically on the quality of the substrate and the homogeneity of the storage material.

For all practical purposes digital systems utilize two methods of storage: nonreturn-to-zero (NRZ) saturation recording, or phase modulation saturation recording. In the first, 0 and 1, are represented by negative and positive saturation of the recording surface, while in the second a double mark is used with 0 being represented by negative saturation followed by positive saturation, and 1 being represented by positive saturation followed by negative saturation. It will be appreciated that the terms positive and negative, 0 and 1, are conventional and that they denote opposites. Fig. 1 compares the two systems on the basis of recording current, storage surface magnetization, and readback voltage. It is assumed that the storage surface employing NRZ recording has been "erased" by uniformly magnetizing the medium to negative saturation (0 state). Information is then inserted by pulsing the recording head so as to change the state of magnetization of the storage surface for a 1. The recording head is left quiescent for a 0. In phase modulation recording, the initial state of the surface is not important since the desired information can be written over whatever digit may be stored. If erasing is required this can be achieved by writing 0's in each cell. The process of interpreting the readback voltage and reconstructing the information to recognizable digital form will be discussed in a later section.

Electrical Design Considerations

ACCESS TIME

The usefulness of a magnetic drum is as great as its ability to perform the required system function of storing a known quantity of data in discrete blocks and to recall them in a given time. The total number of bits N, the number of discrete stores (tracks) n, and the approximate

Table I. Access Time (3% Slip) as Function of Motor Speed and Number of Heads Per Track

Motor			Access Time, Milliseconds		
Supply Frequency, Cps	No. of Poles	Synchronous Speed, Rpm	1 Head Per Track	2 Heads Per Track	4 Heads Per Track
60	2	3,600	17	8.6	4.3
	4	1,800	34	17	8.6
	8	900	69	34	17
	16	450	140	69	34
400	2	24,000	2.6	1.3	0.64
	4	12,000	5.1	2.6	1.3
	8	6,000	10	5.1	2.6
	16	3,000	21	10	5.1
1,000	2	60,000	1.03	0.51	0.26
	4	30,000	2.06	1.03	0.51
	8	15,000	4.12	2.06	1.03
	16	7,500	8.24	4.12	2.06

access time t_a' are defined by system considerations. By the judicious choice of the number of poles and the supply frequency for the drive motor it is usually possible to tailor the system requirements to utilize 60,400 or 1,000 cps (cycles per second) power. Table I shows representative values of access time for drums driven by induction motors utilizing these frequencies. It will be observed that the use of several heads synchronized mechanically to sample a single track can achieve access times as short as 240 μsec (microseconds). As an alternative, the same acess time will be achieved if one head is used and the information is written a number of times around the rotor circumference. Such times, however, are achieved only with a corresponding loss of storage capacity.

Recorded Pulse Length

The total amount of information which can be stored depends on the length of the individual bits recorded on the drum surface and on the rotor diameter. While it is possible to write with a pulse density greater than that naturally associated with the writing head pole gap and the head-to-storage surface separation, this results in a decrease in readback voltage and degeneration of resolution. Karlquist[1] has computed the magnetic field at the surface of the recording medium under stationary conditions. For large values of pulse length (as compared with $d_a'g'$) the exact expression can be approximated as

$$H_x = H_0 d_a' g' \mu' / \pi x^2 \tag{4}$$

where H_x is the magnetizing force at a point on the storage surface in the direction of the storage track, distant x from the center of the recording gap; H_0 is the magnetizing force in the recording gap; g' is the effective recording head pole gap; and μ' is the effective permeability of the storage medium. Writing $x_p = 2x =$ pulse length, and assuming a square-loop recording medium for which the coercive force is H_c at the writing frequency, yields

$$x_p = 2\left(\frac{H_0}{H_c}\,\frac{d_a'g'\mu'}{\pi}\right)^{1/2} \tag{5}$$

Since H_c is difficult to determine under practical writing conditions, a reasonable expedient is to define an arbitrary percentage of the maximum field strength as determining pulse length. If $H_x = 0.2\,H_0$ (20% point) equation 5 can be written as follows:

$$x_p = 2.53(d_a'g'\mu')^{1/2} \tag{6}$$

For an oxide surface, $\mu' \simeq 3$ and for a Ni-Co surface $\mu' \simeq 20$, so that

$$x_p = 4.37(d_a'g'\mu')^{1/2} \quad \text{(oxide)}$$

$$x_p = 11.3(d_a'g')^{1/2} \quad \text{(Ni-Co)} \tag{7}$$

Comparing two surfaces, for the same pulse length

$$d_{a1}'g_1' = (\mu_2'/\mu_1')d_{a2}'g_2' \tag{8}$$

In order to achieve the same packing density on drums of equal size but with different storage surfaces, the recording head pole gap and the head-to-storage surface separation must be decreased in the drum having the greater effective permeability. Alternatively, equal pulse length can be achieved with the same recording head and head-to-storage surface separation if the coercive force of the storage surface is increased, i.e.,

$$H_{c2} = H_{c1}(\mu_2'/\mu_1') \tag{9}$$

It should be remembered that equations 5 through 9 are only as valid as equation 4, which in turn is limited to the domain in which the integral expressions obtained by Karlquist in reference 1 can be replaced by the first term of an asymptotic series. They may be regarded as qualitative indication of the influence of storage surface properties on pulse length.

In a practical system, both d_a' and g' will vary. If $d_a' = d_{a0}(1 \pm K_s)$ and $g' = g_0(1 \pm K_g)$, using equation 6 yields

$$x_p' = 2.53\left(1 \pm \frac{1}{2}K_s \pm \frac{1}{2}K_g\right)(d_{a0}g_0\mu')^{1/2} \tag{10}$$

If "smearing" of the recorded data is to be prevented, the write current pulse period must be sensibly shorter than the time required for the recorded bit to pass under the writing head. This will ensure that the storage surface remains approximately stationary during the writing operation. The rise time of the writing pulse is relatively unimportant under this condition. As a practical convenience the writing current pulse width can be made 20% of this period, i.e.,

$$t_p = 0.2nt_a/N \tag{11}$$

For maximum readback voltage, the recording surface diameter should be capable of accommodating N/n bits per track having maximum length, i.e.,

$$D_0 = Nx_p'/\pi n = 0.8\frac{N}{n}\left(1 \pm \frac{1}{2}K_s \pm \frac{1}{2}K_g\right) \times (d_{a0}g_0\mu')^{1/2} \tag{12}$$

In certain applications, two or more heads are mounted to read information on the same track. If the two heads represent the extremes of head pole-gap variation, if they read what the other has written, and if they are located so that the maximum air-gap variation exists between them at the same time, the difference in recorded pulse length is

$$x_{p1} - x_{p2} = 2.53(K_s + K_g)(d_{a0}g_0\mu')^{1/2} \tag{13}$$

The variation between the leading edges of the two extreme pulses will be one half of the value given in equation 13. Expressed as a fraction of the nominal pulse length, this becomes

$$(x_{p1} - x_{p2})/2x_p = \frac{1}{2}(K_s + K_g) \tag{14}$$

representing a time difference of

$$t_d = \frac{1}{2}(K_s + K_g)nt_a/N \tag{15}$$

which must be taken into account when the use of multiple-headed tracks for fast access applications is considered. If $K_s = 0.1$ and $K_g = 0.2$, it will be seen that the maximum time delay will be 15% of the total pulse period.

Head-to-Storage Surface Separation

In a recent paper, Miyata and Hartel,[2] have shown that, for practical head-to-storage surface separations, the maximum readback voltage can be represented by

$$e_{\max} = C \log_e\left(1 + \frac{d_s}{d_a - a}\right) \tag{16}$$

where $e_{\max}$ = peak readback voltage, C = constant for a particular storage surface and running conditions, d_s = storage surface thickness, d_a = head-to-storage surface separation, and a = recording constant.

The value of the recording constant a varies with the type and thickness of the magnetic surface and with head-to-storage surface separation. For modern oxide coatings and separations of the order of 1 or 2×10^{-3} inches, $a \simeq 0.2$. Similar information is not available for Ni-Co surfaces. It will be assumed, therefore, that the results derived for oxide surfaces can be applied qualitatively to the Ni-Co case.

Rewriting equation 16, differentiating, and putting $K_s = \Delta d_a/d_a$ gives

$$\frac{\Delta e_{\max}}{e_{\max}} = K_s \times \left[\frac{d_a d_s}{(d_a - a)(d_a + d_s - a)}\,\frac{1}{\log_e\left(1 + \dfrac{d_s}{d_a - a}\right)}\right] = \alpha K_s \tag{17}$$

or

$$e_{\max}' = e_{\max}(1 \pm \alpha K_s) \tag{18}$$

The maximum output voltage will be affected in a similar manner by head-setting tolerances, i.e.,

$$e_{\max}' = e_{\max}(1 \pm \alpha K_a) \tag{19}$$

Values of α, computed from equation 17, for representative values of d_a, d_s, and a are as follows:

d_s	d_a	α
0.50	0.50	1.06
	1.00	1.00
	1.50	0.99
	2.00	0.98
	2.50	0.98
0.75	0.50	0.95
	1.00	0.91
	1.50	0.92
	2.00	0.93
	2.50	0.94
1.00	0.50	0.87
	1.00	0.86
	1.50	0.88
	2.00	0.90
	2.50	0.92
1.50	0.50	0.77
	1.00	0.77
	1.50	0.81
	2.00	0.83
	2.50	0.86

The value of a is 0.2 throughout.

HEAD CROSSTALK

The magnetic field of an energized recording head in the plane of the storage surface is approximately that of a small dipole. Other heads in the immediate vicinity will be affected by that component of the field which lies perpendicular to the pole gap faces. In the case of a magnetic drum, all of the heads are oriented in the same direction, i.e., with pole gaps parallel to the axis of rotation. With the use of elementary properties of a small dipole, it can be shown that the component of the writing head field perpendicular to this direction is

$$H_x = \frac{M}{r^3}(1+3\cos^2\theta)^{1/2}\cos\frac{3\theta}{2} \qquad (20)$$

where M is the magnetic moment of the dipole, r is the distance between the dipole center and the receiving point, and θ is the angle between the dipole axis and a line joining the center of the dipole to the receiving point. Equation 20 can be interpreted in terms of the noise voltage induced in a head by another writing in close proximity, i.e.,

$$e_{nc} \propto \frac{e_1}{r^3}(1+3\cos^2\theta)^{1/2}\cos\frac{3\theta}{2} \qquad (21)$$

where e_1 = voltage developed across the writing head by the writing current. Table II compares values of experimentally determined crosstalk with equation 21. It will be observed that agreement is excellent when $\theta=0$ degrees and $\theta=90$ degrees but that the experimental technique shows a zero occurring about $\theta=45$ degrees, while theory predicts that it should occur at $\theta=60$ degrees. This discrepancy may be due to lack of sensitivity in the voltage-measuring equipment and distortion of the simple dipole pattern. Fig. 2 shows experimental values of crosstalk voltage as a function of head separation. The variation from a 3-power law is illustrated by the inclusion of an arbitrary cubic curve.

Table II. Head Crosstalk as a Function of Orientation

Values Compared With a Theoretical Dipole Field

Angle, θ Degrees	Crosstalk: Observed Voltage, $V(\theta)$, Mv	Crosstalk: Observed $\frac{V(\theta)}{V(\theta=0)}$	Crosstalk: Theoretical $\frac{V(\theta)}{V(\theta=0)}$
0	140	1	1
45	0	0	0.234
60			0
90	50	0.357	0.353

Head separation 0.25 inch.

NUMBER OF HEADS

If a proportionality constant K_c is introduced, which may be determined from measurements made when $\theta=0$ degrees, and expression 21 is rewritten in terms of r, it becomes

$$r^3 = \frac{K_c e_1}{e_{nc}}(1+3\cos^2\theta)^{1/2}\cos\frac{3\theta}{2} \qquad (22)$$

Putting $\theta=0$ degrees, $\theta=90$ degrees gives

$$r^3(\theta=0°) = \frac{2K_c e_1}{e_{nc}};\; r^3(\theta=90°) = \frac{K_c e_1}{\sqrt{2}e_{nc}} \qquad (23)$$

i.e., for a given value of e_1/e_{nc}

$$r(\theta=0°) = \sqrt{2}\,r(\theta=90°) \qquad (24)$$

Now, as a good approximation, the effective track width may be considered to be the width of the pole piece d_p together with a width equivalent to the pulse length x_p, i.e.,

$$d_t = d_p + x_p \qquad (25)$$

If the number of heads in any one circuit of the drum is n_h, and the crosstalk contributions of circumferential and axial heads are made equal, then

$$n_h d_t = r(\theta=90) = \frac{1}{\sqrt{2}}r(\theta=0) = \frac{1}{\sqrt{2}} \times \left(\frac{2K_c e_1}{e_{nc}}\right)^{1/3} \qquad (26)$$

Introducing the diameter of the drum yields

$$n_h = \pi D_0/r(\theta=0) = \pi D_0\left(\frac{e_{nc}}{2K_c e_1}\right)^{1/3} \qquad (27)$$

Combining equations 26 and 27 yields

$$n_h^2 = \frac{1}{\sqrt{2}} \cdot \frac{\pi D_0}{d_t} \qquad (28)$$

or

$$n_h = 1.49\sqrt{D_0/d_t} \qquad (29)$$

NOISE

The readback signal is affected by noise due to inhomogeneities and discontinuities in the storage surface. Two major types of noise voltage can be recognized. The first is due to the discrete nature of the surface which gives rise to a typically random noise pattern when viewed on an oscilloscope connected to the read head. The second, which is truly a macroscopic case of the first, is due to relatively large nonmagnetic inclusions which generate single recognizable interfering voltages. In the worst case these may be complete holes in the storage surface. The magnitude of this voltage can be approximated as a percentage of the maximum readback voltage e_{max}. Since a complete void is associated with a sudden change of magnetization from saturation to zero, it is reasonable to expect that this will give rise to a voltage of $1/2e_{max}$. For nonmagnetic inclusions which represent a fraction χ of the total storage surface volume per bit, this voltage will be approximately $1/2\chi e_{max}$. Combining the two effects yields the nominal maximum noise voltage due to the recording surface:

$$e_{ns} = e_{max}\left[\eta(1-\chi)+\frac{1}{2}\chi\right] \qquad (30)$$

where η represents the noise due to microscopic effects. This will be modified by air-gap eccentricity, manufacturing tolerances, and head setting variations so that the effective maximum noise voltage due to surface variations will be

$$e_{ns}' = [1+\alpha(K_s+K_a)](1+K_m) \times \left[\eta(1-\chi)+\frac{1}{2}\chi\right]e_{max} \qquad (31)$$

In addition to noise generated by the magnetic recording surface, the reading head is subject to the influence of other heads which may be writing in its vicinity.

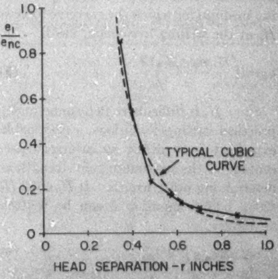


Fig. 2. Typical head crosstalk characteristic

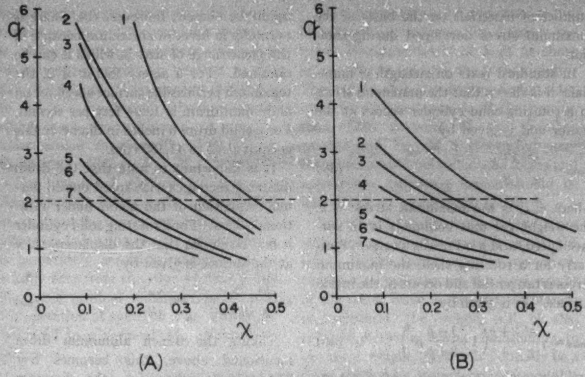


Fig. 3. Variation of signal-to-noise ratio (q) with volume of nonmagnetic inclusion expressed as a fraction of the total storage surface volume per bit (χ)

Curves in A—no interfering writing heads
1—ideal
2—$\alpha=0.91$, $K_a=0.05$, $K_m=0.05$, $K_s=0.05$
3—same as curve 2 except $K_s=0.1$
4—same as curve 2 except $K_s=0.2$
5—$\alpha=0.91$, $K_a=0.2$, $K_m=0.25$, $K_s=0.05$
6—same as curve 5 except $K_s=0.1$
7—same as curve 5 except $K_s=0.2$
Curves in B—one interfering writing head
1—ideal with $(e_1/e_{max})K_c(r\theta)=0.1$
2 through 7—same as curves in A with addition of writing head interference term

For convenience, equation 22 can be written

$$e_{nc}=K_c(r,\theta)e_1 \tag{32}$$

In the general case of a number of heads

$$e_{nc}=e_1[K_c(r_1\theta_1)+K_c(r_2\theta_2)+\ldots] \tag{33}$$

Under normal conditions a certain amount of cancellation between these quantities may exist.

SIGNAL-TO-NOISE RATIO

Just as the noise voltage due to surface inhomogeneities is affected by a number of tolerances, the signal voltage due to recorded data is affected by these quantities also. In computing signal-to-noise ratio, the minimum value of signal should be considered, so that

$$e_{max}'=[1-\alpha(K_s+K_a)](1-K_m)(1-\chi)e_{max} \tag{34}$$

Combining equations 31, 33, and 34 to give signal-to-noise ratio under the worst conditions yields

$$q=e_{max}'/(e_{nc}'+e_{ns}')$$
$$=\frac{[1-\alpha(K_s+K_a)](1-K_m)(1-\chi)}{[1+\alpha(K_s+K_a)](1+K_m)\left[\eta(1-\chi)+\frac{1}{2}\chi\right]+[K_c(r_1\theta_1)+K_c(r_2\theta_2)+\ldots]} \tag{35}$$

For a typical system

$\alpha=0.91$ when $d_a=0.001$ inch; $d_s=0.00075$ inch; $a=0.2$
$K_a=0.05$ for heads set individually
$K_a=0.20$ for heads set collectively
$K_m=0.05$ for selected heads
$K_m=0.25$ for heads taken indiscriminately over the full manufacturing range
$\eta=0.08$
$\frac{e_1}{e_{max}}K_c(r,\theta)=0.1$ for a typical head when $r=0.75$ inch and $\theta=0$

Fig. 3(A) shows the variation of q with χ and K_s for the case when $K_a=0.05$, $K_m=0.05$, and the case when $K_a=0.20$, $K_m=0.25$, when there are no interfering writing heads. The former corresponds to a drum in which selected single heads are set individually, the latter to a drum in which random heads are set collectively by a mechanical expedient. Fig. 3(B) shows the variation of the same quantities in the case when a typical head is writing in proximity to the read head. In both figures the characteristic curves for a mechanically perfect drum are given as reference. If a value of 2 is considered desirable for signal-to-noise ratio, it will be observed that the operating margin is much greater with individually set, selected heads. Indeed, this is essential if simultaneous reading and writing are performed.

READ AMPLIFIER DESIGN

The signals depicted in Fig. 1 are reconstructed into digital information by means of the read amplifier and associated circuits. Typically, the amplifier will consist of a linear amplifier section followed by a nonlinear amplifier-switching section. When several amplifiers are interrogated in parallel, or when it is essential to reproduce the information with an accurate time period, a strobe is usually generated from an accurate clock recorded on the drum surface, and is used to gate the outputs of the read amplifiers. A typical system is shown in Fig. 4(A) and a representative read amplifier design is given in Fig. 4(B). The reconstruction of digital information from the signals shown in Fig. 1 is depicted in Fig. 5.

The important design parameter char-

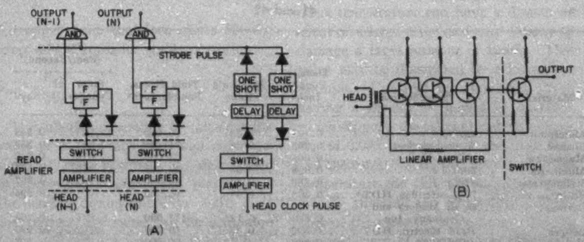


Fig. 4. A—Typical read amplifier and strobe system. B—Typical read amplifier

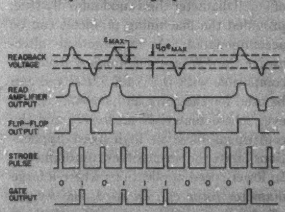


Fig. 5. Reconstruction of NRZ information

acterizing the response of the read amplifier is the input voltage at which full output voltage is generated. Consistent operation can be obtained if this value is chosen so that it is greater than the maximum noise voltage but less than the minimum signal voltage minus the maximum noise voltage, i.e.,

$$e_{nc}'+e_{ns}'<q_0e_{\max}<e_{\max}'-(e_{nc}'+e_{ns}) \qquad (36)$$

where $q_0e_{\max}$ is the threshold voltage of the amplifier. If $q_0e_{\max}$ is designed to be intermediate between these limits, expression 36 reduces to

$$q_0=\frac{e_{\max}'}{2e_{\max}}=\frac{1}{2}[1-\alpha(K_e+K_a)]\times (1-K_m)(1-\chi) \qquad (37)$$

In this case reliable operation can be assured only if the strobe is generated at the peak of the readback signal. This presents serious errors in strobe timing. A better criterion is to design the threshold voltage so that it is midway between the maximum noise signal and some fraction τ of the readback signal less the maximum noise signal. In this case

$$q_0=\frac{1}{2}\tau[1-\alpha(K_e+K_a)](1-K_m)(1-\chi) \qquad (38)$$

Mechanical Design Considerations

In this discussion, interest is limited to a drum comprising a rotor positioned by ball bearings in a stationary shroud which supports salient heads. While many believe that such a drum must run with its axis of rotation vertical, very successful units have been built in which the axis of rotation is horizontal. In either case, the general features of the construction are the same and it should be assumed that the following remarks apply to either unless specific exception is made.

MATERIAL SELECTION

As a first approach to material selection it would seem that, since the bearing rings are normally made from chrome-steel (*SAE 52100*, for example), a number of problems might be solved if the entire unit were made from this material. For drums of small diameter this is undoubtedly true, provided the machining problems can be accommodated. As a practical alternative the *400* series of stainless steels are compatible with the *SAE 52100* and can be used for the rotor and shroud. However, these materials are heavy and the centrifugal force in high-speed applications limits the permissible drum diameter. In lower-speed applications where a large diameter rotor can be used, the weight affects bearing life. In view of these limitations it is instructive to compare a number of materials on the basis of the maximum stress developed during rotation.

In standard texts on strength of materials[3] it is shown that the maximum stress in a rotating solid cylinder occurs at the center and is given by

$$\sigma_{\max}=\gamma\frac{v^2}{g}\,\frac{3+p}{8} \qquad (39)$$

where $\sigma_{\max}$ is the maximum stress, γ is the weight per unit volume, v is the surface speed, and p is Poisson's ratio. Similarly for a rotating shell, the maximum stress is tangential and occurs at the inside surface. It is given by

$$\sigma_{\max}=\gamma\frac{v^2}{g}\,\frac{3+p}{4}\left(1+\frac{1-p}{3+p}\beta^2\right) \qquad (40)$$

where $\beta=a/b$, a is the inner radius of shell, and b is the outer radius of the shell. It will be noted that the maximum stress occurring in a solid cylinder is approximately one half of the stress produced in a hollow shell of the same outside diameter. If σ_{yp} represents the yield point stress and K_{SF} denotes the safety factor employed, equations 39 and 40 may be written

$$v^2=8g\sigma_{yp}/K_{SF}\gamma(3+p)\ \text{(solid)} \qquad (41)$$

$$v^2=4g\sigma_{yp}/K_{SF}\gamma(3+p)\left\{1+\frac{1-p}{3+p}\beta^2\right\}\ \text{(shell)} \qquad (42)$$

If K_{SF} be taken as 2 and $\beta=0.8$, the numerical values of these expressions are compared in Table III on the basis of the square of the linear surface speed. Aluminum can be used to the highest speed, followed by Inconel X and magnesium. Because of its weight (3×Al, 4.5×Mg), and machining properties, Inconel cannot be considered seriously as a general-purpose material for drum construction. In applications in which weight is of paramount importance magnesium is the obvious choice. At lower speeds any of the materials listed might be chosen; however, the choice is normally in favor of aluminum because of the great range of sizes in which it can be obtained. For a safety factor of 2, the maximum permissible surface speed for an aluminum drum is 1,294 feet per second. For a solid drum 6 inches in diameter this is equivalent to 41,180 rpm.

It is important to note that the drum diameter increases with speed due to normal expansion of the metal under rotational stress. For a rotating solid cylinder it can be shown that the displacement u at the surface is given by

$$u=\frac{b^2}{8}\,\frac{2}{1+p}\,(1-p^2)\,\frac{\gamma\omega^2}{gE} \qquad (43)$$

Taking the 6-inch aluminum drum mentioned above, this becomes $u=27.65\times10^{-3}$ inches, i.e., the diameter will increase by some 55 mils when a nominal 6-inch-diameter aluminum drum is rotated at 41,180 rpm. It is essential that any static method of head setting compensate for this growth.

BALANCING

Because of manufacturing tolerances, inhomogeneities in the material, etc., unbalance will be present in a drum rotor and will give rise to radial forces acting on the bearings. The worst case will be that of unbalance occurring at one end of the rotor so that all of the out-of-balance force is applied to one bearing. If bearing considerations limit the maximum permissible radial force to $F_r(\max)$, the maximum permissible out-of-balance moment is

$$M(\max)=\frac{g}{\omega^2}F_r(\max) \qquad (44)$$

If the natural out-of-balance condition of the manufactured rotor is greater than this, it must be balanced to a value less than $M(\max)$. For low-speed applications, it is perfectly satisfactory to subtract weight by drilling the rotor at

Table III. Square of Maximum Surface Speed of Various Materials Computed From Equations 41 and 42

Material	Specification	Density, Pound/Inch³	Poisson's Ratio	Yield Stress, Pounds/Inch²	V²×10⁻⁶, (Feet/Second)² Solid Cylinder	Shell
Aluminum	*2014-T6*	0.101	0.33	60,000	1,903	0.843
Inconel	*X*, cold-worked	0.298	0.33	175,000	1.881	0.834
Magnesium	*AZ80A-T5*	0.065	0.35	34,000	1,675	0.742
Aluminum	*6061-T6*	0.098	0.33	40,000	1.307	0.581
Stainless steel	*440C*	0.280	0.30	60,000	0.692	0.307
Brass	70% cartridge, HDT*	0.308	0.33	64,000		0.294
Gyromet	P. R. Mallory and Company, Inc.	0.620	0.134	114,000	0.626	0.266
Copper	*B188* Electro, HDT	0.322	0.33	50,000		0.220
Stainless steel	*303 (18–8)*	0.290	0.30	30,000	0.334	0.148

* Hard-drawn tubing.

appropriate points; however, high speeds require greater precision and it is desirable to utilize balancing members which can be moved accurately in relation to the axis of rotation. A convenient method is to utilize a number of balancing screws which can be screwed in and out in radial holes similar to conventional gyroscope balancing techniques.

For high-speed applications it is also desirable that balancing be accomplished at or near the operating speed so that adequate balancing can be achieved with a moderately sensitive machine. It is readily seen from equation 44 that if the allowable residual unbalance in a rotor for operation at speed ω_2 is M_2 it must be balanced to a moment M_1, at speed ω_1, given by

$$M_1 = M_2\omega_2^2/\omega_1^2 \tag{45}$$

In the case of a rotor intended for operation at 24,000 rpm, but balanced at 2,000 rpm, the implication is that the sensitivity of the balancing machine must be 144 times better than would be necessary if balancing were done at the operating speed. Under normal conditions such an increase in sensitivity is not practical.

MECHANICAL STABILITY

Suppose a force F_B, applied at the center of the rotor, deflects the center of gravity a distance d from the axis of rotation when the rotor is supported at the bearings; then, for small displacements

$$F_B = K_r d \tag{46}$$

where K_r will be called the coefficient of rigidity of the rotor. If the rotor is rotating, and has a total weight W_r, this deflection will produce a centrifugal force. The rotor will be stable or unstable as

$$K_r > \frac{W_r}{g}\omega^2 \quad \text{(stable)} \tag{47}$$

or

$$K_r < \frac{W_r}{g}\omega^2 \quad \text{(unstable)}$$

An upper limit therefore exists for the speed of the rotor for stable operation:

$$\omega_{\max}^2 = \frac{K_r g}{W_r} \tag{48}$$

It is almost impossible to calculate the value of K_r for any particular rotor design; it can, however, be determined experimentally. Since the majority of the bending occurs where the shafts meet the spinner, its value can be estimated fairly well, once a few models have been tested.

As a consequence of equation 48, in high-speed applications, the distance between bearings should be kept as short as possible to enhance the rotor rigidity. This means that the driving motor rotor must be attached outside of the bearing span. The presence of this member cantilevered beyond the bearing may introduce additional unbalance forces and appropriate balancing means should be introduced to compensate for them.

The size of the bearing selected for use will depend on the orientation of the rotor. If the axis of rotation is vertical, one bearing will support the total weight of the rotor, while if the axis of rotation is horizontal, the weight will be distributed between them. In addition to the rotor weight there will normally be a force applied by springs or other suitable means to the rotor in the direction of the axis of rotation to stabilize the balls in the bearings. This "preload" force will act equally on both bearings in either system. The choice of actual bearing size is best left to the recommendation of the bearing manufacturers since specific operating conditions may require special consideration. Experience has shown that their recommendations in regard to bore alignment and bore tolerances should be followed accurately for reliable service and long life.

THERMAL STABILITY

An important aspect of drum design is thermal mismatch between members constructed of different materials. This is particularly true with the advent of solid-state computers. Previously, air conditioning was provided to keep the operating temperatures low enough to achieve proper cooling of solid-state components in the presence of vacuum-tube circuitry. As a secondary effect, drums were operated in a stable temperature environment. However, with the elimination of vacuum tubes by solid-state devices, air conditioning is unnecessary and the drum must operate under natural conditions. Sudden changes in environment temperature can have a disastrous effect on a large drum and may destroy or damage a large number of heads. This point may be illustrated by a numerical example. Consider an aluminum drum having a shroud 24 inches in diameter which supports the salient head with an air gap of 0.001 inch between the head face and the rotor. A sudden drop in external temperature can chill the shroud without greatly affecting the temperature of the bulk of the rotor for some time. Taking the coefficient of expansion of aluminum as 13×10^{-6} inches per inch per degree Fahrenheit (F), a sudden change of 6.4 F in the temperature of the shroud will cause the heads to touch the storage surface.

If aluminum is chosen as the rotor material a mismatch in thermal coefficients of expansion occurs between the aluminum shaft and the chrome-steel inner race of the bearing and between the chrome-steel outer race and the aluminum housing. For a moderate rise in temperature, bearings with a large radial clearance should be chosen so that there is adequate margin for expansion of the inner race by the expansion of the aluminum shaft. The stress produced can be computed from expression 5:

$$\tau_{\max} = \delta c^2 \left\{ \frac{b}{E_r}\left(\frac{b^2+c^2}{c^2-b^2} + p_r\right) + \frac{b}{E_s}(1-p_s) \right\}^{-1} \tag{49}$$

where $\tau_{\max}$ is the maximum shear stress, δ is the magnitude of the interference fit between the inside diameter of the bearing race and the outside diameter of the shaft, b is the outside diameter of the shaft, c is the inside diameter of the bearing races, E_s, E_r are the values of Young's modulus for the shaft and the race, and P_s, P_r are the values of Poisson's ratio. The shear stress should be kept below the manufacturer's recommended value.

Conclusions

It should not be assumed that successful drum construction is solely a simple application of the principles enumerated. As with any electromechanical device, much depends on the skilled hands of the mechanic making and fitting the parts, as well as on the experience of the designer. Attention to fine detail cannot be overstressed. The "roundness" of the bearing spots, the fit of the outer race in the bores, the quality of the finish grind on the rotor, the precision of balancing, the relief of the machining stresses after boring the shroud, are all examples of items as essential to success as the proper choice of bearings or the right size of driving motor.

Viewed simply as a precision mechanical device, a memory drum is an impressive unit in which a highly machined, precise rotor rotates at several thousand rpm within 0.001±0.0001 inch from as many as 500 separate heads. From a chemical viewpoint, the preparation of the magnetic surface poses many problems, particularly in the case of plated deposits in which subtle amounts of addition agents and the temperature of deposition may have a profound effect on the magnetic properties of the deposit.

Lest the task appear too difficult, it should be said that each of these difficulties can be overcome by employing special techniques. Machining a rotor at running speed to correct for deformations due to the dynamic forces present is one example. The discovery of broader range additives for the control of the plating bath is another. As with any complex technical process, tight control at every step is the secret of success, with as much detailed attention being given to the measurement of quality as to the production of the item itself.

A well-made, reliable magnetic drum is capable of many years of faultless service at a cost per bit which gives it premier place as a medium-speed high-capacity storage device. At the upper end of the storage spectrum it is flanked by ferrite cores, cryogenic devices, and thin films; at the lower end it yields to disk memories and magnetic tapes. In applications of moderately fast data rate (up to 500 kc) and medium access time (in milliseconds), the magnetic drum has established its utility and will be a part of the computer designer's repertoire for a long time to come.

Nomenclature

a = recording constant: inner radius of shell
b = outer radius of shell: shaft diameter
B_0 = nominal magnetic induction in storage medium
B' = effective magnetic induction in storage medium
c = inside diameter of bearing race
C = constant
d = deflection
d_{a0} = nominal air gap
d_a' = effective air gap (also d_{a1}', d_{a2}')
d_p = width of recording head pole piece
d_s = thickness of storage surface
d_t = width of storage track
D_0 = nominal rotor diameter
D' = effective rotor diameter
e_{max} = nominal peak voltage at readback head
e_{max}' = effective peak voltage at readback head
e_{nc} = peak voltage due to crosstalk between heads
e_{ns} = peak voltage due to magnetic surface noise
E_r = Young's modulus for bearing race
E_s = Young's modulus for shaft
F = centrifugal force
F_B = bending force
$F_r(\max)$ = maximum permissible radial force
g = acceleration due to gravity
g_0 = nominal recording head pole gap
g' = effective recording head pole gap (also g_1', g_2')
H_0 = maximum magnetizing force in recording gap
H_c = coercive force of storage medium (also H_{c1}, H_{c2})
H_x = magnetizing force at point distant x along storage track from center of pole gap
I_m = intensity of magnetization in storage surface
K_a = factor defining variation in head readback voltage due to head-setting variations
K_c = crosstalk coefficient; also $K_c(r_1\theta_1)$, $K_c(r_2\theta_2)$
K_e = factor defining air-gap variation (rotor eccentricity)
K_g = factor defining variation of recording head pole gap
K_m = factor defining variation of recording head output voltage due to manufacturing tolerances
K_r = coefficient of rigidity
K_s = factor defining variation of magnetic properties of storage surface
K_{SF} = safety factor
M = magnetic moment of dipole equivalent of recording head: unbalanced moment; also $M(\max)$
n = number of recording tracks
n_h = number of heads in one complete turn around drum
N = total number of bits stored
P = Poisson's ratio (also P_r, P_s)
q = minimum signal per maximum noise
q_0 = read amplifier threshold level
r = radius: distance
t_a = access time
t_a' = approximate access time
t_d = maximum time delay (skew) between pulses in register tracks
t_p = duration of writing current pulse
u = displacement of rotor surface due to centrifugal force
v = surface velocity
W = unbalanced weight
x_p = recorded pulse length (also x_{p1}, x_{p2})
x_p' = effective pulse length
$\beta = a/b$
γ = weight per unit volume
δ = interference fit between shaft and bearing race
η = noise voltage due to microscopic effects
θ = angle
μ' = effective permeability of storage surface (also μ_1', μ_2')
σ_{max} = maximum stress
σ_{YP} = yield point stress
τ = maximum shear stress
χ = volume of nonmagnetic inclusion as a fraction of total storage surface volume per bit
ω = radial velocity (also ω_1, ω_2)

References

1. CALCULATION OF THE MAGNETIC FIELD IN THE FERROMAGNETIC LAYER OF A MAGNETIC DRUM, **O. Karlquist.** *Transactions*, Royal Institute of Technology, Stockholm, Sweden, no. 86, 1954.

2. THE RECORDING AND REPRODUCTION OF SIGNALS ON MAGNETIC MEDIUM USING SATURATION TYPE RECORDING, **J. J. Miyata, R. R. Hartel.** *Transactions*, Professional Group on Electronic Computers, Institute of Radio Engineers, New York, N. Y., vol. PGEC-8, no. 2, 1959, pp. 159–69.

3. STRENGTH OF MATERIALS—VOL. II (book), **S. Timoshenko.** D. Van Nostrand Company, Inc., Princeton, N. J., second edition, 1947, pp. 245–49.

4. *Ibid.*, p. 246.

5. *Ibid.*, pp. 241–42.

A New Function Generator

C. W. EGGERS
NONMEMBER AIEE

T. W. SZE
MEMBER AIEE

IN ANALOG COMPUTING technology as well as in the laboratory study of control instrumentation or simulation systems, it is frequently desirable to be able to provide a given arbitrary periodic function as a forcing function. However, no device will reproduce exactly every arbitrary function which might be required. For this reason, considerable effort has been applied toward the development of devices which will reproduce an approximation to a wide variety of functions. The term, "function generator" has been given to devices of this type.

The existing function generators may be separated into the following:

1. Analytic function generator. These generators employ the property that a certain number of analytic functions are the known solutions of differential equations. For example, a sinusoidal function may be generated by setting up an ordinary differential equation using three operational amplifiers.[1]

2. Diode function generator. This is the most widely used type in present-day analog computing facilities. In these devices, an input variable, x, produces an output, $F(x)$, which consists of a number of straight line segments between break points, which may be fixed or variable. The change in slopes is accomplished by biasing a number of diodes, each to one of the break points.

3. Tapped potentiometers. If a potentiometer is rotated with the angular position proportional to a variable, x, the wiper

---

Paper 60-1203, recommended by the AIEE Computing Devices Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Chicago, Ill., October 9–14, 1960. Manuscript submitted June 7, 1960; made available for printing August 3, 1960.

C. W. EGGERS and T. W. SZE are members of the faculty of the Dept. of Electrical Engineering, University of Pittsburgh, Pittsburgh, Pa.

This work is partially based on an *M. S. Thesis* by C. W. Eggers at the University of Pittsburgh.

voltage will vary approximately linearly with x between the voltages applied to the two ends of the potentiometer. In the tapped potentiometer used for function generation, the voltages at a number of taps along the potentiometer are forced to specified values and, as the wiper is rotated between any two points, $F(x)$ will vary linearly between the two corresponding voltages. In this manner, a curve of $f(x)$ versus x may be approximated by straight line segments, and functions of time may be obtained by generating x proportional to time.

4. Follower-type function generator. An example of this is the photoformer, in which a mask is placed on a cathode-ray tube and a servosystem with a photosensitive pickup forces the trace to follow the mask.

5. Function generation by synthesizing harmonics.[3] Since any periodic function may be written as the sum of a finite or infinite number of harmonics, an arbitrary periodic function may be synthesized by producing and combining a number of harmonics with proper amplitude and phase relationships.

A new function generator has been developed, constructed, and tested successfully by the authors, and results will be presented in this paper. This function generator employs the combined design techniques of digital circuits and analog computing circuits.

Development of the New Function Generator

Theory of Operation

The problem to be considered is the generation of an approximation, $F(t)$, to a function of time, $f(t)$, which is valid for $0 \leq t < T$. The function may be periodic with period T, or it may be aperiodic where $0 \leq t < T$ is the region of interest.

The following scheme of approximation is used in developing the new function generator. The period T is divided into equal intervals. In each interval, the function is approximated, either by a constant value, by a constant slope, or by a curve having some higher derivative constant. A commutating device is developed and used to generate constant voltages in each interval. The constant voltages are then integrated the required number of times to produce the approximation.

In the approximation process the period T is first divided into n equal intervals in an analytic or graphical representation of the function. Within each interval, say the ith interval, the approximation of the function may be written as

$$F_i(t) = K_i\left[t - \frac{(i-1)T}{n}\right]^p + F_{i-1}\left[\frac{(i-1)T}{n}\right]$$

for

$$\frac{(i-1)T}{n} \leq t < \frac{iT}{n} \tag{1}$$

and

$$F_i(t) = 0$$

for

$$t < \frac{(i-1)T}{n}$$

or

$$t \geq \frac{iT}{n}$$

with

$$F_0 = f(0)$$

where p may be any positive integer including zero. Finally,

$$f(t) \approx F(t) = \sum_{i=1}^{n} F_i(t) \tag{2}$$

The exponent, p, and the set of constants, K_i, are chosen for the best approximation to the function, $f(t)$, and these choices form the mechanics of the setup procedure. The exponent, p, must first be chosen on the basis of the general form of $f(t)$ and on the desired degree of accuracy of approximation for a given number of interval divisions. For example, if the function consists of a series of steps, the choice is $p=0$ so that the approximation will consist of a series of steps of magnitudes K_i. If a good approximation of function consists of a series of segments of constant slopes, the case of $p=1$ applies. The approximation is then made by straight line segments, and

$$F_i(t) = K_i\left[t - \frac{(i-1)T}{n}\right] + F_{i-1}\left[\frac{(i-1)T}{n}\right] \tag{3}$$

for

$$\frac{(i-1)T}{n} \leq t < \frac{iT}{n}$$

and the constants, K_i, are obtained as the slopes of F_i in the ith interval. That is,

$$K_i = \frac{dF_i}{dt} \tag{4}$$

The result of this set of calculations is the formation of a staircase function

$$c(t) = \sum_{i=1}^{n} c_i(t) \tag{5}$$

where

$c_i(t) = aK_i = C_i$ within the ith interval

and

$c_i(t) = 0$ in all other intervals (6)

and a is a constant scale factor. The steps in the formation of this function are shown in Fig. 1.

For the cases of $p \geq 2$, the approximation of $f(t)$ is made by segments of parabolic arcs (for $p=2$), or of higher order. Procedure for $p \geq 2$ is similar to that for $p=0$ or $p=1$ in principle, but the actual process of finding K_i for each interval becomes quite difficult except for an analytic function, instead of an arbitrary function.

Implementation of the Function Generator

The theory just presented has involved the mathematical procedures for the approximation of a function. To generate a known but arbitrary time function, one will first calculate $c_i(t)$ for each interval, equations 6. The different $c_i(t)$ are sampled, one at a time and in sequence, by a commutating device to form $c(t)$ or equation 5. Then finally, this function $c(t)$ is integrated p times, and multiplied by the scale factor a^{-1} to produce the approximation

$$F(t) = a^{-1} \int \int \ldots \int c(t)\,dt\,dt \ldots dt$$

and

$$F(t) \approx f(t) \tag{7}$$

These steps are indicated by the block diagram of Fig. 2. The potentiometers in Fig. 2 supply the voltages C_i, which are sampled in sequence by the commutating device. The integration may be per-

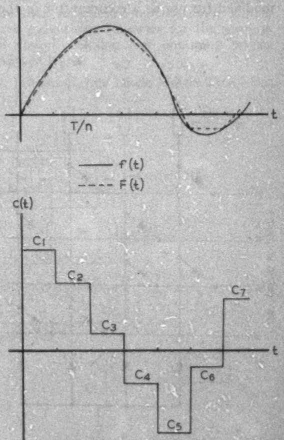


Fig. 1. Process of approximation and formation of staircase function

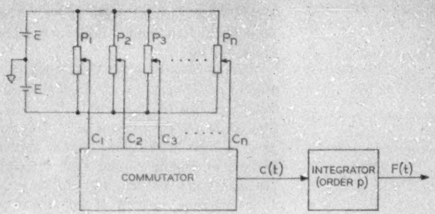


Fig. 2. Block diagram of a function generator using a commutating device

formed by an operational amplifier appropriately connected as an integrator.

The function generator may be made periodic by making the commutating device periodic. On the other hand, if an aperiodic operation is desired, the commutating device is designed to sample the c_i's only once.

Commutating Diode Matrix Tree

An essential element of the new function generator is the commutating device. Commutating devices used in function generators and in other applications have been reported previously.[2-9] The function generator presented in this paper employs a modified diode matrix as a commutating device.

The diode matrix and its variations (diode pyramid and diode rectangle) are well known to the digital circuit designers.[10] Such circuits have been widely used in the design of combinational logic networks such as the encoding and the decoding networks in a digital computing system. The basic structure of a diode matrix is modified for use as a commutator in the development of the new function generator.

The commutator actually used in the function generator is a 4-input diode matrix tree capable of commutating 16 segments. Its principle of operation, however, may be more easily explained by using a 2-input diode matrix, capable of commutating only 4 segments. Fig. 3 is a basic diode matrix for use as a commutator. This may be developed from a static diode matrix by the following modifications:

1. The switching voltages applied to the terminals A, A', B, B' have waveforms as shown in Fig. 4(a). Notice that the voltage waveform applied at B or B' has a frequency of twice that at A or A'. Furthermore, the voltages at A and A' and at B and B' are logically complements of each other. Therefore, these voltages are the outputs of a binary counter, with a count pulse train having the same frequency as B or B'.

2. The load resistors, instead of being connected to a common constant voltage as in the normal operation of a diode matrix, are individually connected to the wipers of potentiometers as shown in Fig. 2. Examples of the waveforms of voltages at these points are shown in Fig. 4(b).

3. The outputs of the diode matrix: e_1, e_2, e_3, e_4, are mixed by an OR gate to produce $c(t)$.

The principle of commutation is based on two facts of the diode logic circuits (assume ideal diodes):

1. In a diode AND logic circuit, the output voltage is equal to *the most negative one* among the input voltages and the supply voltage, which are either zero or positive.

2. In a diode OR logic circuit, the output voltage is equal to *the most positive one* among the input voltages, which are either zero or positive.

Based on the former, the waveforms of e_1, e_2, e_3, e_4, which are really outputs of AND gates, can be drawn as shown in Fig. 4(c). Based on the latter, $c(t)$ is then produced by mixing e_1, e_2, e_3, and e_4. This is equivalent to saying that the commutating diode matrix has sampled the c_i in sequence for $i=1,2,3,4$, etc. If the switching voltages are periodic, the sampling will also be periodic.

To commutate and sample 2^k segments, a k-input commutating diode matrix which contains $(k+1)2^k$ diodes is required. Here is seen a basic weakness of a diode matrix, namely that the number of diodes rises sharply as k increases. Variations of the diode matrix, such as the diode pyramid and the diode rectangle[10] have been developed with the number of diodes

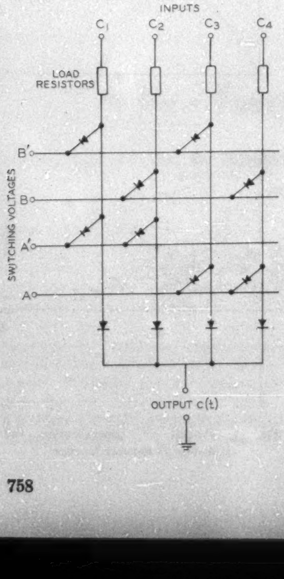


Fig. 3 (left). A 4-segment commutating diode matrix

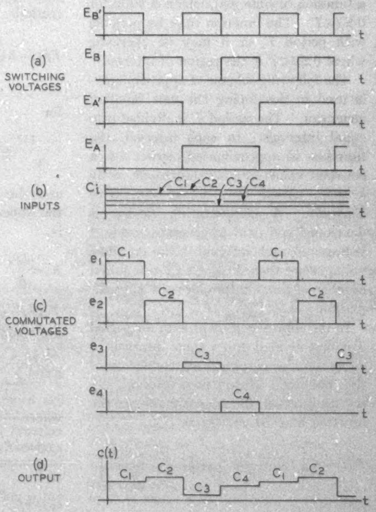


Fig. 4 (right). Principle of the commutating diode matrix

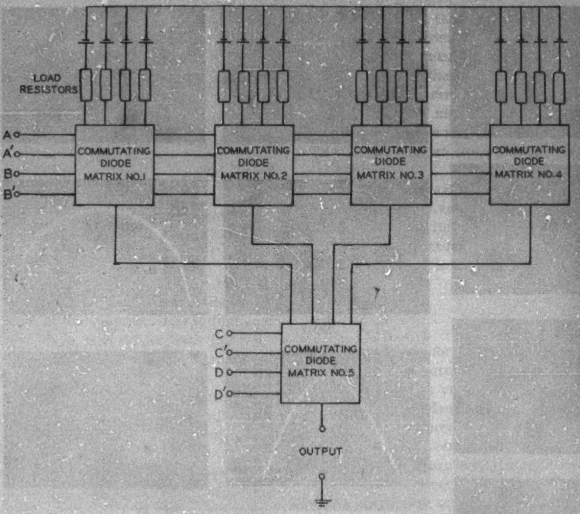

LOAD RESISTORS
A
A'
B
B'
COMMUTATING DIODE MATRIX NO.1
COMMUTATING DIODE MATRIX NO.2
COMMUTATING DIODE MATRIX NO.3
COMMUTATING DIODE MATRIX NO.4
C
C'
D
D'
COMMUTATING DIODE MATRIX NO.5
OUTPUT

Fig. 5. A commutating diode matrix tree for commutating 16 segments

Fig. 6. Laboratory model of the function generator

A—Front view
B—Rear view

used substantially reduced. The function generator presented in this paper uses a variation of the diode matrix which is more economical than a diode pyramid or a diode rectangle of the same commutating and sampling capacity. This variation is referred to here as a commutating diode matrix tree, and is shown in Fig. 5 for 16 segments or $k=4$. In this scheme the 16 segments are divided into 4 groups of 4 segments each. A commutating diode matrix is used for each group, and a fifth commutating matrix is used to commutate from group to group. The economy of such a scheme is shown in Table I in which the number of diodes is compared. Note that the number of diodes refer to the *commutating* diode matrix rather than the conventional *static* diode matrix.

Table I. Comparison of Number of Diodes Required for Variations of Commutating Diode Matrix

k	Number of Segments (2^k)	Number of Diodes for Commutation: Matrix	Pyramid	Rectangle	Tree
1	2	4	*	*	*
2	4	12	*	*	*
3	8	32	*	*	28
4	16	80	72	64	60
5	32	192	152	128	124
6	64	448	312	256	252

* Same as that required for a matrix.

Laboratory Model

A laboratory bench model of the function generator based on the principle presented here was designed, built, and tested successfully at the University of Pittsburgh. This generator was designed as an additional equipment for the existing analog computing facilities.

Design Considerations of the Laboratory Model

Fig. 6 shows a front and rear view of he laboratory model. Fig. 7 shows the block diagram for the function generator. Some of the design considerations will be described briefly as follows:

1. Timing Circuit. The laboratory model was built for a function approximation using 16 segments. To supply the switching voltages for the commutating diode matrix tree, a 4-bit binary counter, with a trigger flip-flop supplying the count pulses, is provided. The counter outputs are then separated from the diode matrix tree by cathode followers for the purpose of circuit isolation and obtaining proper voltage levels.

2. Commutating Diode Matrix Tree. For

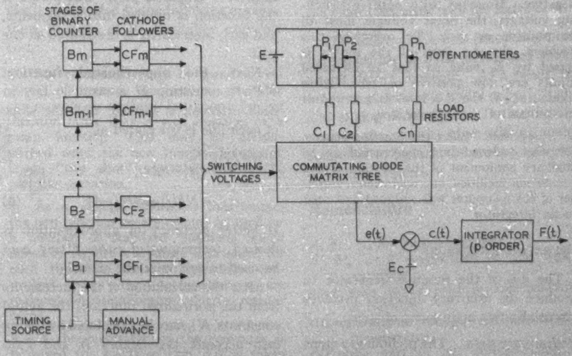

STAGES OF BINARY COUNTER
CATHODE FOLLOWERS
Bm
CFm
Bm-1
CFm-1
B2
CF2
B1
CF1
TIMING SOURCE
MANUAL ADVANCE
SWITCHING VOLTAGES
E
P1
P2
Pn
POTENTIOMETERS
LOAD RESISTORS
C1
C2
Cn
COMMUTATING DIODE MATRIX TREE
e(t)
c(t)
Ec
INTEGRATOR (p ORDER)
F(t)

Fig. 7. Block diagram of the new function generator

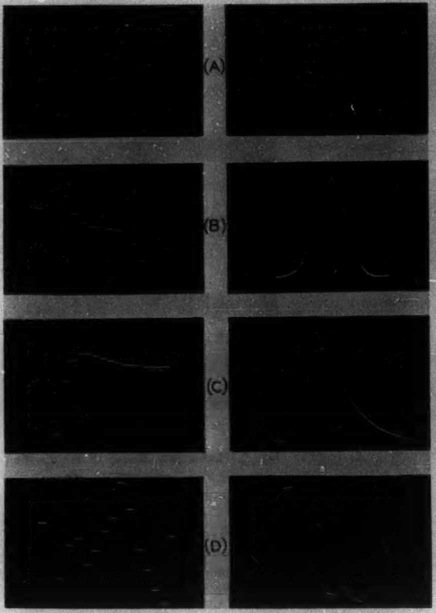


Fig. 8. Examples of the function generator outputs

- A—c(t) and F(t) of a sine wave
- B—c(t) and F(t) of a Gaussian function
- C—c(t) and F(t) of te^{-at}
- D—c(t) and F(t) of an arbitrary time function

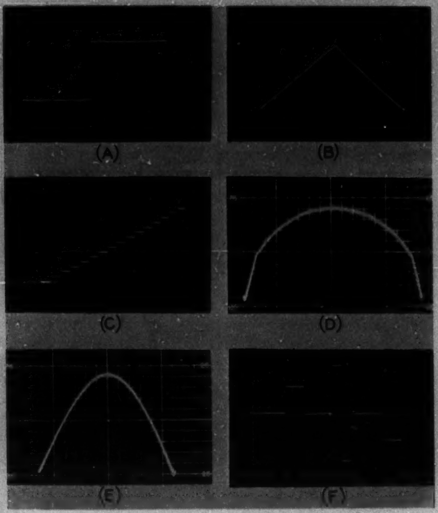


Fig. 9. More examples of the function generator outputs

- A—Step function
- B—Triangular wave
- C—Staircase or jump function
- D—Semicircular function
- E—Rectified sine wave
- F—An arbitrary discrete function

Vol.
Pt.
196

commutating 16 segments, the commutating matrix used is the one shown previously in Fig. 5. The input voltages are taken from the wipers of the potentiometers. It should be noted that these input voltages correspond to the slopes of the approximated function for the case $p=1$. It follows that these voltages may be positive or negative. However, with positive switching voltages, the input voltages must all be positive or zero. To overcome this problem, the d-c level of the inputs is raised by E_c volts so that all increased inputs are now positive or zero. After commutation, the d-c level is restored at the output of the commutator.

3. Integrator. Since the function generator was designed and constructed as an auxiliary equipment of the existing analog computing facilities, one operational amplifier of the computer was assigned and used as an integrator.

OPERATING PROCEDURE

The use of the function generator to produce an arbitrary function involves essentially two main series of steps:

Preliminary steps. The preliminary steps are mainly mathematical steps taken to determine the settings of the controls of the function generator. As a first step, it may be convenient to consider a normalized function $Nf(x)$, where N is a constant and x is an independent variable. This function may be either of tabulated form or a graphical plot. One period of the function is divided into n segments, and each interval Δx is designated as one unit.

Next, the approximated function, $NF(x)$, consisting of a straight line in each interval, is formed, and the slope of each line is calculated as

$$\frac{\Delta[NF_i(x)]}{\Delta x}=\Delta[NF_i(x)]=NF(i)-NF(i-1) \tag{8}$$

If better accuracy of approximation is desired, techniques of curve fitting may be used to minimize the error.[11,12]

After the calculation of the increments from the normalized function, the actual constants, K_i, may be calculated by taking into account the factors N and T/n. Thus,

$$K_i=\frac{\Delta F_i(t)}{\Delta t}=\frac{\Delta[NF_i(x)]}{NT/n} \tag{9}$$

The scale factor, a, is obtained next. It may be conveniently set to the scale factor of the integrator so that $a=RC$, where R and C are the resistance and the capacitance respectively, used in a conventional connection of an operational amplifier as an integrator. The choice of R-C must, in turn, depend upon the possible range of C_i, and the necessary range of K_i. The range of C_i is set by that of the diode matrix which has some limiting outputs, E_{max} and E_{min} respectively, so that

$$|E_{max}-E_{min}|=|C_{imax}-C_{imin}|\geq RC|K_{imax}-K_{imin}| \tag{10}$$

The range indicated by equation 10 also determines the magnitude of E_c. Although this voltage may be set in any manner so long as the diode matrix inputs become all positive, E_c is preset here to allow equal range of C_i of both signs. It follows that

$$|E_c|=\frac{1}{2}(E_{max}+E_{min}) \tag{11}$$

so that the range of C_i may be rewritten

$$-\frac{1}{2}(E_{\max}-E_{\min}) \leq C_i \leq \frac{1}{2}(E_{\max}-E_{\min}) \tag{12}$$

With these values fixed, equations 5 and 6 are determined for a given arbitrary function. Now, it may be written that the diode matrix output for any interval is

$$E_i = C_i + |E_c| \tag{13}$$

where E_c is the artificial d-c level change and is restored before the integrator stage, see Fig. 7.

Set-up steps. The mechanical steps of setting up the generator are quite simple. They are essentially the adjustment of the potentiometers so that equation 13 is satisfied for each interval i. The counter is provided with a manual advance so that the inputs may be sampled one at a time at the leisure of the operator. This allows the voltage for each interval to be measured by the precision instrument available on the analog computer, and has the advantage of bypassing some of the effects of the nonideal components in the diode matrix. Alternately, the negative voltage E_e may be combined ahead of the voltage measurement which is then made to satisfy equation 6 and no calculation of E_i is necessary. When the potentiometer adjustments are completed, the function generator output should be the approximated function.

TEST RESULTS

The oscillograms in Figs. 8 and 9 are samples from the very large group of functions that the function generator is capable of generating. In each case of Fig. 8, a staircase waveform of slopes, $c(t)$, and the generator output, $F(t)$, are shown. Fig. 9 shows some other waveforms generated by the function generator.

DISCUSSION OF ACCURACY

The experimental model of the function generator was constructed using only commercial, nonprecision, and inexpensive components. For example, all 60 diodes required for the commutating diode matrix tree are *1N34A*, one of the most inexpensive types available, 39 cents each. With such low-cost components, concern may arise as to the accuracy attainable by the use of the function generator.

Two possible sources of error exist in the generation of an approximation to an arbitrary function. The first is the approximation itself while the second is the implementation of the approximation. When, in the set-up procedure, the voltages, C_i, are precisely measured, the characteristics of the diodes and the potentiometers become unimportant and and errors of implementation are comparable to those in the analog computing equipments. The errors of the first type are, for the same number of segments, comparable to those in any function generator employing an approximation by straight line.

To study the magnitude of the error involved, a sine wave was approximated so that the desired function is $f(t)=\sin wt$ while the generated function is $F(t)$. The per-cent error is defined here as

$$\frac{\int_0^T |f(t)-F(t)|\,dt}{\int_0^T |f(t)|\,dt}$$

and computation for the sine wave example yields an error of 1.29 per cent. For comparison, a calculation based on the definition of Hamer[12] yields a maximum per-cent difference between $f(t)$ and $F(t)$ as 1.93 per cent. Finally, actual measurement of harmonic distortion was made with the generator generating the approximated sine wave. When a single integration was used ($p=1$), the generator output shows a harmonic distortion of 4.07 per cent. When a double integration was used ($p=2$), the generator output shows a harmonic distortion of 1.24 per cent. These figures of errors include errors of both types.

Summary

The primary purpose of the function generator is to provide an approximation to a specified function as a forcing function to an actual or a simulated system. Consequently, the function generater developed here is used as a part of existing analog computing facilities. However, aside from this primary application, the new function generator can be particularly well-suited to the following applications because of its flexibility and low cost:

1. For computer study of transient response of systems, the function generator can provide an arbitrary but known forcing function other than a step function or a ramp function. This may be accomplished when the new function generator is on "one shot" operation instead of periodic operation.

2. As a college laboratory instrument, the new function generator may be used for a number of standard functions which normally require a separate instrument for each function. This is especially convenient when the potentiometers are replaced by precalibrated plug-in resistor cards.

3. If the integrator of the generator is bypassed (or $p=0$) so that $c(t)$ is taken as the output of the generator, the function generator becomes well suited for the studies of discrete signal systems. Several examples are:

   (*a*). With all inputs but one set to zero, the output is a pulse of width T/n and an amplitude determined by a single input. If a time varying signal is applied to that input, the amplitude of that pulse will be modulated by the signal. The modulating signal could be obtained from a second function generator. This will allow the extension of studies on pulse amplitude modulation beyond the use of a single frequency sine wave.

   (*b*). If all potentiometers are set to nonzero values, the voltages, $e(t)$ and $c(t)$, are of the same form as the output of a holding circuit in a sampled-data control system. Thus the function generator may be used in a computer study to simulate the output of a holding circuit.

   (*c*). The function $c(t)$ may be so adjusted that it can be used to represent the forcing function of a difference equation. The use of this function and this function generator would allow the simulation of certain digital systems on an analog computer.

4. If there are two such function generators available at the computer, they may be designed to allow for an adjustable delay between their time scales. Calculation of correlation functions, which involves delay, may be easily carried out with such analog computing equipment.

5. The concept of a matched signal or a matched filter suggests that, for optimum freedom from noise, the signal component of the input be $h(-t)$ for a system with an impulse response of $h(t)$.[13] With the new function generator, it is of no difficulty to generate $h(-t)$ once the waveform of $h(t)$ is known and the time scale is shifted.

6. The function generator may also be used in simulating environment in certain equipment tests. One such example is the radar test in which a vehicle at various ranges, velocities, and accelerations is to be simulated. One instrument that is used for such testing[14] provides for an input from a function generator to simulate a maneuver. The use of the function generator should add to the versatility of the radar tests made with the instrument.

The design of the new function generator makes use of both digital and analog circuit design concepts. As a result, the function generator is useful in both digital and analog simulations. It thus further expands the capacity of an existing analog computer to include the study of discrete or digital systems.

Nomenclature

$A, B, \ldots, X$ = logical variables
$A', B', \ldots, X'$ = negations of the logical variables
a = a scale factor used in approximating a function
$c(t)$ = a variable voltage, the output of a commutator
$c_i(t)$ = the part of $c(t)$ corresponding to interval i

C_i = a constant voltage associated with the interval i
$e(t)$ = a variable voltage, the output of the diode matrix
e_i = a variable voltage associated with the interval i
E = a constant power supply voltage
E_c = a change of d-c voltage level
$f(t)$ = a time function to be generated by the function generator
$F(t)$ = the time function generated by the function generator
i = an integer, the number of a segment or an interval along the time axis
K_i = a constant, equal to the derivative of $F_i(t)$
N = a number, a normalizing scale factor used in approximating a function
n = an integer, the number of segments to be commutated by the commutator
P_i = a potentiometer associated with interval i
t = time variable
T = the period of a periodic function
x = a dependent variable

References

1. Analog Computer Techniques (book), C. L. Johnson. McGraw-Hill Book Company, Inc., New York, N. Y., 1956.

2. The Design of Function Generators Using Short-Time Memory Device and Nonlinear Elements, A. W. Revay, D. J. Ford. *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 77, May 1958, pp. 143–52.

3. Electronic Analog Computers (book), G. A. Korn, T. M. Korn. McGraw-Hill Book Company, Inc., 1952, pp. 254–55.

4. A Universal Unit for the Electrical Differential Analyzer, R. Tomovich. *Journal*, Franklin Institute, Philadelphia, Pa., vol. 254, 1952, pp. 143–51.

5. A Versatile Electronic Function Generator, R. Tomovich. *Ibid.*, vol. 257, 1954, pp. 109–20.

6. Function Generation by Integration of Steps, W. Comley. *WESCON Convention Record*, Institute of Radio Engineers, New York, N. Y., pt. 4, 1957, pp. 279–83.

7. Pulse and Digital Circuits (book), J. Millman, H. Taub. McGraw-Hill Book Company, Inc., 1956, pp. 339–43.

8. Telemetering—Commutating Devices, A. A. McKenzie, H. A. Manovgian. *Electronics*, New York, N. Y., Apr. 1956, pp. 173–74.

9. Electronic Commutators by Means of Minority Carrier Deflection in Semiconductors, G. C. Szikiai, L. R. Hill. "Proceedings, Symposium on the Role of Solid State Phenomena in Electric Circuits," Brooklyn Polytechnic Institute, New York, N. Y., Apr. 23–25, 1957, pp. 175–78.

10. Synthesis of Electronic Computing and Control Circuits, Staff of Harvard Computing Laboratory. Harvard University Press, Cambridge, Mass., 1951, chap. X.

11. Approximation Errors in Diode Function Generator, N. Ream. *Journal of Electronics and Control*, New York, N. Y., vol. 7, no. 1, July 1959, pp. 83–96.

12. Optimum Linear-Segment Function Generation, Howard Hamer. *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 75, Nov. 1956, pp. 518–20.

13. Electronic Circuits Signals and Systems (book), S. J. Mason, H. J. Zimmerman. John Wiley & Sons, Inc., New York, N. Y., 1960.

14. Radar Testing for a War Environment, R. W. Hanford. *National Convention Record*, Institute of Radio Engineers, pt. 5, 1959, pp. 208–17.

Current-Operated Diode Logic Gates

HENRY REINECKE, JR.
NONMEMBER AIEE

THE LITERATURE contains many articles on the design of voltage-operated diode logic gates. Perhaps the major disadvantage of this type of gate is the restriction of the design to two or three stages which does not, in general, allow realization of minimal networks or which alternately requires the insertion of a gain device to re-establish the voltage levels. This paper formulates the general realization procedures applicable to the synthesis of current-operated diode logic networks. The current-operated gate is essentially the dual of the voltage-operated gate; this relationship is established to substantiate the utilization of existing synthesis methods. Two distinctly different realization procedures are described, both equally general, and the choice is usually dictated by the nature of the specific transmission function. The Boolean operations of addition and multiplication (commonly referred to as OR and AND respectively) are more conveniently realized with the use of the "dual method." The operation of sum modulo two (or function selection) is more economically synthesized using the "lattice method." For a further discussion of the terms see the section on add and multiply, reference 1.

The operation of complementation assumes a unique meaning in current-operated systems because the reversal of a current may be accomplished by simply transposing the two lines connecting the gate output and the load.

A static multiplier capable of simultaneously multiplying two 4-bit binary numbers was designed employing current-operated techniques exclusively. This particular type of multiplier was selected because it advantageously utilizes both realization methods and requires the synthesis of a wide variety of transmission gates.

Preliminary Derivational Steps

Conventions and Dual Relationships

The realization procedures for current-operated diode logic gates may be derived from the rules employed for voltage-operated gates, using the principles of duality. Table I lists the pertinent standard dual relationships and Table II summarizes the conventions applied to voltage-operated gates and the corresponding conventions for the current-operated gates.

In accordance with these conventions, the 2-input add and multiply (alternately referred to as OR and AND respectively) gates and their respective duals are shown in Fig. 1. For background information on duality and the construction of dual graphs see references 2 and 3. The polarity of the current sources is established by employing the convention; when the dashed line connecting points a and b, which intersects the voltage source, is traversed in the direction from a to b and the voltage source is polarized to the right (or the normal direction of current flow

Table I. Standard Dual Relationships

Voltage	Current
Resistance	Conductance
Series	Parallel
Loop	Node
Short circuit	Open circuit
Voltage source	Current source
Diode conducting	Diode nonconducting

Table II. Conventions Applied to the Gates

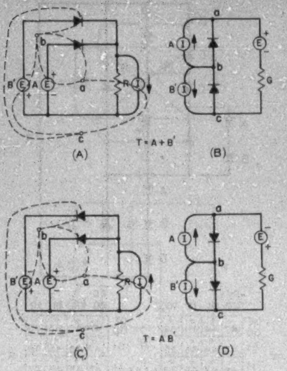
Parameter	Voltage-Operated	Current-Operated
Logic level	$\pm E$	$\pm I$
State 1	$+E$	$+I$
State zero	$-E$	$-I$
Source impedance	low	high
Load impedance	high	low
Connection of stage inputs	in parallel	in series
Connection of stages	in series	in parallel

Paper 60-1188, recommended by the AIEE Computing Devices Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Chicago, Ill., October 9–14, 1960. Manuscript submitted June 15, 1960; made available for printing August 2, 1960.

Henry Reinecke, Jr., is with Non-Linear Systems, Inc., Del Mar, Calif.

The author expresses his appreciation to Prof. S. H. Caldwell for his supervision of the related thesis and to Kenneth H. Perry and Dr. Irwin L. Lebow for their valuable comments on the presentation of the material. Thanks are due also to Daniel H. Galvin for his help in the construction of the static multiplier and to Alberta L. Samara and Nancy E. Stebbins for their assistance in typing the manuscripts. This paper is based in part upon a thesis submitted by the author to the Electrical Engineering Department of the Massachusetts Institute of Technology in partial fulfillment of the requirements for the degree of Master of Science.

Vol.
Pt. 1
196



INPUTS SHOWN FOR STATE A=B=1

Fig. 1. Basic add and multiply gates

is clockwise), the current source in the dual circuit is confluent on *a*. Conversely, if the voltage source is polarized in the opposite direction the corresponding current source is confluent on *b*. Similar conventions apply in determining the polarity of the diodes. Traversing the dashed line from *a* to *b*, crossing the diodes, if the preferred direction of conduction is to the right (or clockwise), the diode in the dual circuit points toward *a*. If the preferred direction of conduction is to the left (or counterclockwise) the diode then points toward *b*.

In practice, the load resistor associated with the add gate is generally returned to a voltage more negative than the magnitude of the voltage used to represent a low (or 0) input; the current source shunting the load resistor in Fig. 1(A) is merely the Thévenin equivalent of this situation. The converse is true for the multiply gate shown in Fig. 1(C), and hence the current source is reversed.

Basic Add and Multiply Gates

All diodes shown in Fig. 1 are ideal and the inputs are shown for the state $A=B=1$. The voltage source in the current-operated gates serves to reverse-bias all diodes not forward-biased by their associated current sources. Also, the load conductance is high, and proportional to the load resistance of the voltage-operated gate. Determination of the transmission of the gates is readily accomplished by considering the four possible combinations of the inputs. The add circuit is redrawn for all four input states and a reasonable set of network parameters is shown in Fig. 2; idealized results are included. In addition, the various mesh currents and their associated paths are shown to indicate the direction of the load current. In general, the current paths are obvious by inspection, although some elaboration on Fig. 2(B) may be beneficial. Both current sources appear to be in the direction to forward-bias the diodes. However, the voltage drop across the load is 0.2 volt which subtracts from the 1.0-volt source. The remaining 0.8 volt is sufficient to reverse-bias the diodes and direct the current as shown.

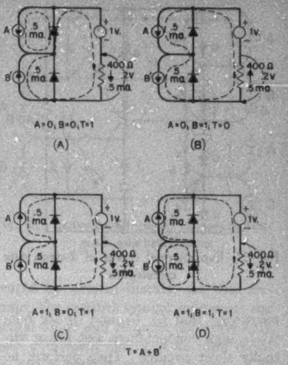
Using the convention that current down through the load represents a 1 (one), we may deduce that the transmission is represented by $T=A+B'$. The above procedure may be repeated for the multiply gate which has a transmission of $T=AB'$. A few preliminary conventions may now be stated: 1. Diodes in series pointing up represent elements of add gates. 2. Diodes in series pointing down represent elements of multiply gates. 3. If the input current source points up when the variable is in state 1, the input is represented by the unprimed literal; conversely, if the input current source points down when the variable is in state 1, the input is represented by the primed literal. 4. Current down through the load represents a transmission of 1.

Directing our attention to the diodes at this point, which in the nonideal situation will exhibit a voltage drop of a few tenths of a volt when conducting, and observing that this voltage would be of the same polarity as the auxiliary voltage source, it appears that the series bias-voltage source might be eliminated.

Realization and Design of Current-Operated Gates

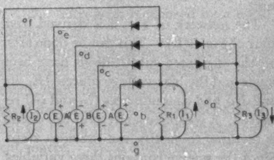
Dual Method of Realization

The dual method of realizing current-operated diode logic networks utilizes to advantage the diode voltage drop. The general principles of realization are established by analyzing the simple multistage network of Fig. 3 which represents the transmission function $T=AB+A'C$. The networks of Fig. 3(A) and Fig. 3(B) are dual and use ideal diodes. As suggested above, if the ideal diodes were replaced by nonideal diodes it might be possible to eliminate the three bias-voltage sources shown in Fig. 3(B). Using an ideal diode, in series with a voltage source as the equivalent circuit of a nonideal diode, accounts for the forward drop of real diodes and may suffice as a first-order approximation. The characteristics of a germanium junction diode and the proposed equivalent circuit are shown in Fig. 4 for comparison. It was previously

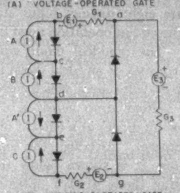


T=A+B'

Fig. 2. Current-operated add gate shown for the four input states

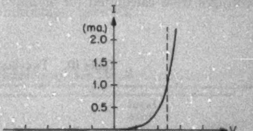


(A) VOLTAGE-OPERATED GATE

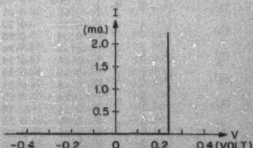


(B) CURRENT-OPERATED GATE
INPUTS SHOWN FOR STATE A=B=C=1
T=AB+A'C

Fig. 3. Two-stage diode logic gates



(A) GERMANIUM DIODE TYPE CTP-521



(B) IDEAL DIODE IN SERIES WITH 0.24V SOURCE

Fig. 4. Voltage-current characteristics

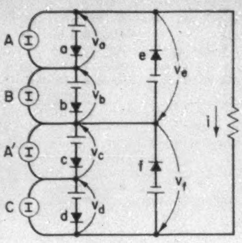
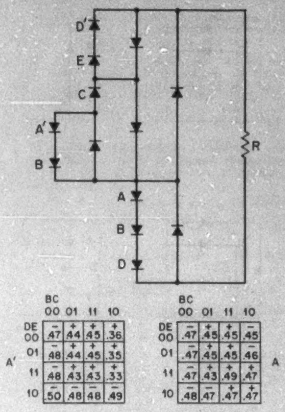


Fig. 5. Experimental results of 2-stage gate calculations

mentioned that G_1, G_2, and G_3 are relatively high conductances and hence virtually short circuits. They perform no logical function and may be eliminated but actually it is desirable to retain G_3 to provide means of sensing the load current.

The network of Fig. 5 is derived from that of Fig. 3(B). This can be done after all the suggested substitutions have been made and the bias-voltage sources (E_1, E_2, and E_3) and conductances (G_1 and G_2) removed. A complete analysis of this circuit is required to justify the previous steps. Since the system contains three variables, there are eight distinct input states for which the circuit operation must be determined. The nonideal diode equivalent circuit consists of an ideal diode in series with a 0.24-volt source, which corresponds to the voltage drop across a *CTP-521* (germanium diode) at a current of 1.0 ma (milliampere). Although the logic level is 0.5 ma, it is shown below that any conducting diode always operates at a current equal to twice the logic level. The current sources are typically 0.5 ma and the load resistance is equal to 40 ohms. Using these values, the various voltages were calculated (or assumed, in the cases where the values are known to be indeterminate, by considering all similar elements identical), and tabulated on the diagram. The bottom values represent the measured data and are included to facilitate comparison of the results. The arrowheads indicate the reference directions of the branch voltages.

Referring to Fig. 5 and Table III(a), the first input state to be considered is $A=B=C=0$, and the polarity of the current sources correspond to the convention that the source points up when the literal is in state 1. Current source A' points up, which causes diode c to conduct, and allows the current to circulate in this mesh. The second mesh current of interest is described by a path linking current sources A and B, conducting diode c, and current source C, and is completed via the load. Diodes a, b, d, e, and f remain essentially nonconducting because of the bias applied across their associated branches. Kirchoff's voltage law requires that the algebraic sum of the voltages around any closed loop must equal zero. To meet this requirement the terminal voltages of sources A, B, and C must assume the values indicated. The voltage at the node joining diodes a and b is indeterminate, but if the diodes are assumed to be identical, the voltage across each would be the same. Ideally, the load current would be 0.50 ma; the reduced magnitude of the measured value is due to the deviation of the actual diode characteristics from that of the assumed equivalent circuit. The transmission is zero because the output current is directed up through the load.



DE \ BC	00	01	11	10
00	− .47	+ .44	+ .45	+ .36
01	− .48	+ .44	+ .45	+ .35
11	− .48	+ .43	+ .43	+ .33
10	− .50	− .48	− .48	− .49

DE \ BC	00	01	11	10
00	− .47	+ .45	+ .45	− .45
01	− .47	+ .45	+ .45	− .46
11	− .47	+ .43	+ .49	+ .47
10	− .48	− .47	+ .47	+ .47

$$T = A'BD' + A'BE + CD' + CE + ABD$$
$$= (D'+E)(A'B+C) + ABD$$

Fig. 6. Multistage gate and measured load currents

Table III. Two-Stage Gate Calculations, See Fig. 5

Line Designation	Input State A	B	C	v_a, Volts	v_b, Volts	v_c, Volts	v_d, Volts	v_e, Volts	v_f, Volts	i, Ma	Output State, T
(a)	0	0	0	−0.005 −0.05	−0.005 −0.01	+0.24 +0.24	−0.25 −0.19	+0.01 +0.06	+0.01 −0.04	−0.50 −0.49	0
(b)	0	0	1	−0.12 −0.08	−0.12 −0.16	+0.13 +0.13	+0.13 +0.13	+0.24 +0.24	−0.26 −0.26	+0.50 +0.44	1
(c)	0	1	0	−0.25 −0.27	+0.24 +0.24	+0.24 +0.24	−0.25 −0.21	+0.01 +0.04	+0.01 −0.02	−0.50 −0.49	0
(d)	0	1	1	−0.48 −0.48	+0.24 +0.24	+0.13 +0.13	+0.13 +0.13	+0.24 +0.24	−0.26 −0.26	+0.50 +0.44	1
(e)	1	0	0	+0.24 +0.24	−0.25 −0.18	−0.005 +0.02	−0.005 −0.00	+0.01 −0.05	+0.01 +0.07	−0.50 −0.49	0
(f)	1	0	1	+0.24 +0.25	−0.25 −0.20	−0.25 −0.28	+0.24 +0.24	+0.01 −0.03	+0.01 +0.05	−0.50 −0.49	0
(g)	1	1	0	+0.13 +0.13	+0.13 +0.12	−0.12 −0.11	−0.12 −0.12	−0.26 −0.25	+0.24 +0.23	+0.50 +0.43	1
(h)	1	1	1	+0.13 +0.13	+0.13 +0.13	−0.48 −0.47	+0.24 +0.24	−0.26 −0.26	+0.24 +0.24	+0.50 +0.44	0

Circuit operation for the remaining input states will be discussed more briefly. In particular, to eliminate extensive repetition, the meshes will be described by indicating the sources and/or diodes which they include, using only the letter associated with each, and ordering the letters to correspond to the direction of current flow. The two mesh currents in Table III (b) are *A-B-e* (this abbreviated form is used to indicate that the current from source *A* flows through source *B* and finally returns via diode *e*) and *C-A′-e*-load. Diodes *a*, *b*, and *f* are obviously nonconducting, but diodes *c* and *d* shunt sources *A′* and *C* respectively. These diodes would ideally be nonconducting but, owing to the bias present, a small amount of current will flow. The effect of such shunt paths is to reduce the actual output current. The transmission is 1 because the output current is down through the load.

Three meshes are established in Table III(c). The two minor ones are: *B-b* and *A′-c*; the one including the load is *A-b-c-C*-load. The transmission is zero and no significant "shunt paths" exist for this configuration. The three meshes present in (d) are: *B-b*, *A-b-e* and *C-A′-e*-load and diodes *c* and *d* shunt a fraction of the current in a manner similar to that discussed for (b). However, the direction of the current in the load shows the transmission to be 1 as required. Operation of circuits of (e) through (h) are very similar to the previous circuits as may be determined by comparing (e) and (a), (f) and (c), (g) and (b), and finally (h) and (d). In general, the function of the upper and lower halves of the circuits has been interchanged but the configuration of the meshes having circulating currents is the

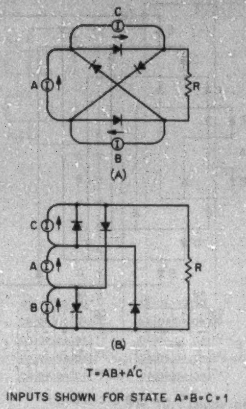


Fig. 7. Generalized current-operated lattice network

same. For completeness, the conducting meshes are tabulated below:

Table III

Line Designation	Conducting Meshes
(e)	*A-a*; *a-B-A'-C*-load
(f)	*A-a*; *C-d*; *a-B-A'-d*-load
(g)	*A'-C-f*; *f-B-A*-load
(h)	*C-d*; *A'-d-f*; *f-B-A*-load

In general, the calculated voltages are in excellent agreement with the measured values, except in the cases where the voltages were indeterminate. Since the voltage across the 40-ohm load (a ±0.50 millimeter) never exceeds 0.02 volt, it is a sufficiently good approximation to a short circuit with respect to the 0.24-volt drop across a conducting diode.

Reviewing the results of Fig. 5 and Table III, it is concluded that the auxiliary series bias-voltage sources and the intermediate conductances may be eliminated. Furthermore, this realization demonstrates the method of combining stages, and points out the similarity between multistage voltage-operated and current-operated gates. The two terms of the minimum-sum function are separately realized with the use of multiply gates. The term AB is obtained by applying current sources A and B to two diodes in series pointing down. Similarly, the term $A'C$ is formed by applying current sources A' and C to two diodes in series pointing down. The output of these gates is a current which has a specified direction depending on the state of the inputs. Addition of the two terms is accomplished by a 2-input add gate which requires two series-connected diodes pointing up. The outputs of the two multiply gates serve as inputs to the add gate which completes the realization.

The operation of complementation has a unique meaning in current-operated systems. Clearly, if the output lines of the network in Fig. 3(B) were transposed without changing the position of the load, the current would flow through the load in the opposite direction. Retaining the original convention that current down through the load represents a transmission of 1, it would be concluded that the transmission of the gate was the complement of the realization intended. Another approach to this problem is the realization of the complementary transmission function. If this is carried out the resulting network will be identical to that of Fig. 3(B) revolved 180 degrees about a horizontal center line. This reorientation of the circuit reverses the direction of the sources and the diodes, and interchanges the add and multiply gates, but the network remains unaltered. Hence, the realization of any specified transmission function is unique but indistinguishable from its complement. This property results from the fact that the output current is not referred to a specified reference current, as is the output voltage in voltage-operated systems.

The 4-stage circuit of Fig. 6 illustrates the application of the design techniques and the method of interconnecting the diodes to accomplish "logical" addition and multiplication. All current sources have been omitted, the appropriate convention being that the implicit current source points up when the associated literal is in state 1; conversely, the source points down when the associated literal is in state zero. Hence, for the state $A=B=C=D=E=1$, all current sources are directed upward except those represented by primed literals. In this diagram, the rectifier symbol represents the actual diode and replaces the previously used equivalent circuit of an ideal diode in series with a voltage source. The given transmission function is realized by considering the more subordinate portions first, which in this case corresponds to $A'B$. This term is realized by a basic multiply gate.

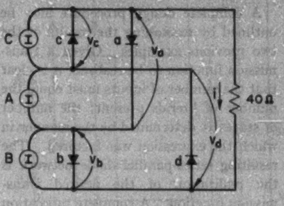


Fig. 8. Experimental results of lattice network gate calculations

Each remaining term is subsequently synthesized and the outputs to these stages may be employed as inputs to succeeding stages in a manner identical to that for which the input is a primary current source. A ±0.50-ma center-reading milliameter, having an impedance of approximately 40 ohms, was used to make the measurements and it also constituted the output load resistance. The measured output currents for the 32 input states are tabulated with the use of Karnaugh maps[4]; a positive current represents a transmission of 1 and a negative current represents a zero transmission. The minimum-sum form of the transmission equation may be written directly from the maps as $T=A'BDD'+A'BE+CD'+CE+ABD$, which is the desired result. A 2-stage realization would require 18 diodes, in contrast to the 13 diodes used in the multistage network.

Table IV. Lattice Network Gate Calculations, See Fig. 8

Line Designation	Input State A	B	C	v_a, Volts	v_b, Volts	v_c, Volts	v_d, Volts	i, Ma	Output State, T
(a)	0	0	0	−0.01 −0.04	−0.01 +0.02	+0.01 +0.01	+0.01 +0.01	−0.50 −0.50	0
(b)	0	0	1	+0.01 +0.03	+0.01 −0.01	−0.26 −0.26	+0.24 +0.24	+0.50 +0.50	1
(c)	0	1	0	−0.26 −0.26	+0.24 +0.24	+0.01 +0.04	+0.01 −0.02	−0.50 −0.50	0
(d)	0	1	1	−0.22 −0.22	+0.24 +0.24	−0.26 −0.26	+0.24 +0.24	+0.50 +0.50	1
(e)	1	0	0	+0.24 +0.24	−0.26 −0.26	+0.24 +0.24	−0.22 −0.22	−0.50 −0.50	0
(f)	1	0	1	+0.24 +0.24	−0.26 −0.26	+0.01 −0.03	+0.01 +0.05	−0.50 −0.50	0
(g)	1	1	0	+0.01 −0.01	+0.01 +0.03	+0.24 +0.24	−0.26 −0.26	+0.50 +0.50	1
(h)	1	1	1	+0.01 +0.04	+0.01 −0.02	−0.01 −0.01	−0.01 −0.01	+0.50 +0.50	1

A complete design procedure may be outlined by examining the results of the two previous examples. Given a transmission function to be realized, it is clear that the number of inputs must equal the number of literals present; the number of stages is determined by the manner in which the expression was factored. The resulting series-parallel diode network is the realization of the desired transmission function. A complete tabulation of the design procedure is included in Appendix I for convenient reference.

LATTICE METHOD OF REALIZATION

The lattice method of realizing current-operated diode logic networks uses four diodes connected in series to form a ring as the basic building block. The unique transmission characteristics of this configuration will be discussed for the generalized case. A 4-diode lattice network is shown in Fig. 7(A) and is redrawn in (B) to resemble more closely the previous diagrams and to facilitate the application of the established conventions to this method. All rectifier symbols represent nonideal diodes and the load impedance is low, effectively a short circuit, as was the case for the previous method.

The lattice structure is a general 3-input network and its transmission is represented by $T=AB+A'C$. All conducting meshes for the eight input states may be determined from the tabulated data of Fig. 8 and Table IV, and are listed below:

Table IV

Line Designation	**Conducting Meshes**
(a)	*C-A-B*-load
(b)	*A-B-d, d-C*-load
(c)	*B-b; C-A-b*-load
(d)	*B-b; A-b-d; d-C*-load
(e)	*C-c; A-c-a; a-B*-load
(f)	*A-C-a; a-B*-load
(g)	*C-c; B-A-c*-load
(h)	*B-A-C*-load

It is interesting to note that diodes *a* and *c* remain nonconducting for the four input states when current source *A* is in state zero; and that diodes *b* and *d* remain nonconducting for the four input states when current source *A* is in state 1. Inspection of the output state column of Fig. 8 and Table IV, reveals the fact that the realization satisfies the desired transmission function.

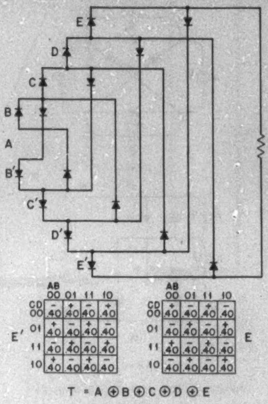
The principle of operation parallels that of networks realized with the use of the dual method, where it was observed that current circulates in various meshes, thereby establishing the direction of the current in the mesh which includes the load; current down through the load signifies a transmission of 1. All conducting diodes are observed to link two meshes and the current in each is equal to twice the logic level. The eight measured values of output current are all equal to 0.50 ma, and were obtained using the ±0.50-ma center-reading milliameter.

The design rules depend primarily on the relative directions of the current sources with respect to each other. For the general lattice transmission function, $T=AB+A'C$; either of two conventions may be chosen because of the symmetry of the network. Referring to Fig. 7(B), if current source *A* points up when in state 1, the two product terms will be *AB* and *A'C* respectively, if current sources *B* and *C* point up when their inputs are in state 1. Addition of the terms is inherent in the lattice network which realizes the entire transmission function.

Complementation of any of the variables contained in the function requires reversal of its associated current source. To complement the transmission, the entire network may be revolved 180 degrees about a horizontal center line or the output lines may be transposed. As noted in the discussion of the dual method, the realization is unique and identical to that of the complemented function.

The design procedures for realizing lattice networks are summarized in Appendix I. Comparing this lattice realization with the dual realization of Fig. 3(B) for the same transmission function, a marked reduction may be observed in the number of diodes required. This economical feature is one of the primary advantages of the lattice network. One obvious difference in this method is that current source *A* is not associated with a specific diode, and it performs basically a function of path selection. Substituting a current source B' for the current source *C* yields a transmission of $T=AB+A'B'=A\oplus B'$. This function represents the operation of sum modulo 2 (alternately referred to as ring sum, circle sum, or "exclusive or") and finds many useful applications in digital systems.

Multistage gates may be realized by cascading a number of lattice networks, or by employing the outputs of lattice networks as the inputs to another lattice network (i.e., replacing current sources *B* and *C* in Fig. 7 by the outputs of lattice networks). A more general form of the transmission equation is $T=T_aT_b+T'_aT_c$, where the subscripts correspond to the input positions designated by the capital letters in the previous expression. This expression indicates that the inputs are not restricted to simple current sources; they include currents derived from gate outputs. The operation of sum modulo 2 on five literals is accomplished by cascading four lattice configurations as shown in Fig. 9.



CD \ AB	00	01	11	10
00	− .40	+ .40	− .40	+ .40
01	+ .40	− .40	+ .40	− .40
11	− .40	+ .40	− .40	+ .40
10	+ .40	− .40	+ .40	− .40

CD \ AB	00	01	11	10
00	+ .40	− .40	+ .40	− .40
01	− .40	+ .40	− .40	+ .40
11	+ .40	− .40	+ .40	− .40
10	− .40	+ .40	− .40	+ .40

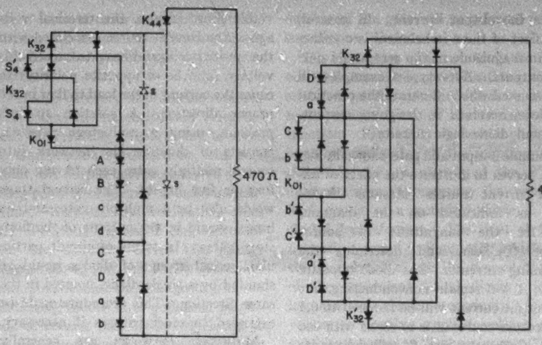
Fig. 9. Multistage lattice networks and measured load currents

The measured output currents are tabulated using two Karnaugh maps from which the transmission function may be directly determined as $T=A\oplus B\oplus C\oplus D\oplus E$. The magnitude of the output current for all 32 input states is equal, which is apparently a result of the symmetry of the network. A 2-stage voltage-operated gate would require 96 diodes, in contrast to the 16 used in the cascaded lattice realization.

HYBRID AND MULTIPLE OUTPUT CIRCUITS

In general, any arbitrary transmission function may be realized using either the dual or the lattice method, although the number of diodes required may be different. Therefore, in the interest of economy, it may be advantageous to consider hybrid realizations which result in more minimal networks. The standard minimizing techniques currently in use reduce the transmission function to its minimum-sum or minimum-product form. This is satisfactory for the realization of 2-stage gates which are extensively used in voltage-operated systems. No practical procedure has been devised which will always result in the simplest transmission expression of any order, where order is synonymous with the number of stages required in the realization. In addition, it would be advantageous to have a method of minimization based on the three binary operations add (+), multiply (·), and sum modulo 2

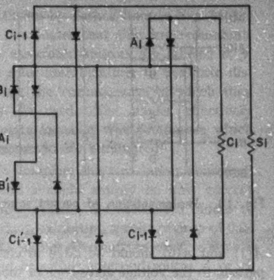
Vol.
Pt. 1
1960



$K_{42} = K'_{44}[K_{01}(K_{32} \oplus S_4)' + K_{32}(K_{32} \oplus S_4) + CDab + ABcd]$
(A)

$S_4 = K_{01} \oplus Cb \oplus Da \oplus K_{32}$
(B)

Fig. 10. Diode logic gates. A—K_{42}. B—S_4



$C_i = C_{i-1}(A_i \oplus B_i) + A_i(A_i \oplus B_i)'$
$S_i = C'_{i-1}(A_i \oplus B_i) + C_{i-1}(A_i \oplus B_i)'$

Fig. 11. Multiple output realization

(⊕). Curtis describes a procedure based on this approach.[5]

Two examples of hybrid realizations extracted from the series of carry and sum gates used in the static multiplier are included. The transmission function of Fig. 10(A) is realized with the dual method employed, and the circuit uses as one of its inputs the output of two cascaded lattice networks. Remarks describing the significance of the diodes, drawn in dashed lines, are contained in the following section. The network of of Fig. 10(B) is essentially a cascaded lattice structure whose inputs represent dual realizations of the product terms represented by Cb and Da.

A single example of a multiple output realization will be presented to clarify the definition and to demonstrate the design technique. The particular pair of transmissions chosen for this illustration are the sum and carry functions associated with a parallel binary adder. If A_i and B_i are the addend and augend bits in the ith binary position, the sum S_i and carry C_i may be expressed as follows, using C_{i-1} as the carry from the adjacent column toward the least-significant-bit end of the word.

$$\begin{aligned} S_i &= A_i \oplus B_i \oplus C_{i-1} \\ &= C_{i-1}'(A_i \oplus B_i) + C_{i-1}(A_i \oplus B_i)' \\ C_i &= A_iB_i + A_iC_{i-1} + B_iC_{i-1} \end{aligned}$$

Expanding in terms of sum modulo 2 yields

$$\begin{aligned} C_i &= A_i(A_i'B_i' + A_iB_i) + C_{i-1}(A_i'Bl + \\ &\qquad A_iB_i') \\ &= A_i(A_i \oplus B_i)' + C_{i-1}(A_i \oplus B_i) \end{aligned}$$

The sum modulo 2 of A_i and B_i, and its complement, are both present in the carry and the sum; therefore we may realize this portion of the functions separately and use it as a common input. A simple lattice network having this transmission is shown in Fig. 11, with its output applied to the inputs of two other lattice networks. Straightforward realization produces the sum and carry as shown. Since the first lattice is required to supply current to two lattices in parallel, the logic levels in the first lattice (i.e., the magnitude of current sources A_i, B_i and B_i') must be doubled at all three inputs. This output current then divides equally, producing the proper output levels for the sum and carry networks. The usefulness of this arrangement is enhanced when we evaluate the relative economies of this realization as opposed to the 2-stage voltage-operated gates. For the latter, a total of 25 diodes are required; 16 diodes for the sum and 9 for the carry. A saving in excess of 50% is achieved with the use of the multiple output lattice network consisting of 12 diodes.

PRACTICAL DESIGN CONSIDERATIONS

There are a number of practical problems associated with the design of current-operated diode logic gates. These include modified realization techniques in special cases, requirements imposed on auxiliary circuits, and general restrictions present in any nonideal system. A pertinent question arises concerning multiple inputs which are dependent on a common variable, where "multiple inputs" refers to either the number of times a specific literal and its complement are repeated for a single gate realization or throughout the entire system. In voltage-operated equipment, all inputs requiring the same literal (or function) are connected to a common point which may be a flip-flop or gate output. Existing digital components usually exhibit low output impedance characteristics and are capable of driving many inputs. The requirements for current-operated systems are markedly different. Components are needed which have a number of high-impedance outputs. These outputs should preferably be isolated, and identical in their performance. Neither devices nor circuits which meet all these specifications are presently available, although a reasonable approximation may be derived with the use of standard components.

Using a conventional flip-flop or a differential amplifier as shown in Fig. 12, both the output and its complement may be paired and connected to all inputs dependent on this binary variable. Conversion of this voltage into many identical current sources is accomplished within the gate circuits in the manner illustrated in Fig. 13. Since the impedance presented to any input current source is relatively low, either a single conducting diode or a series circuit of conducting diodes and current sources, a good approximation to a current input is achieved by inserting series resistance between the network and the $A(1)$ and $A(0)$ voltage inputs. Clearly, the potential difference between

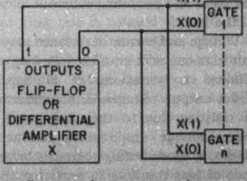


Fig. 12. System wiring scheme

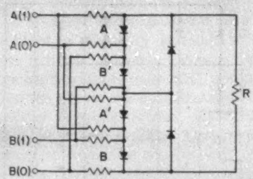


Fig. 13. Implementation of current sources

$A(1)$ and $A(0)$ should be large in comparison to the voltage across one or more conducting diodes which constitute the current source load.

Examining Fig. 13 reveals that both A and its complement are required as inputs to the network. Current source A is derived by inserting a resistor between $A(1)$ and the anode of the diode, and an equal resistor between $A(0)$ and the cathode. The complemented current source is obtained by interchanging the connections to $A(1)$ and $A(0)$, effectively inverting the source and reversing the current with respect to current source A. Inputs B and B' are obtained in a similar manner. It should be pointed out that the load on any flip-flop depends only on the number of inputs it must supply, and is independent of the state of the flip-flop. The load reflected by the parallel resistor-diode branches is essentially constant and balanced; the state of the input merely determines the direction of the current.

Another important factor is the determination of the magnitude of the current to be used to represent the logic levels. This problem is slightly more involved. Actual diode volt-ampere characteristics must be considered because severe current and voltage imbalances may result in marginal circuit performance. This item will be more fully discussed below. In equipment constructed to date flip-flops and current buffers were employed having outputs of ±5 volts, and providing a voltage difference of 10 volts. All series resistors were 10 kilohms with the result that the logic level was 0.50 ma. This set of values proved to be very satisfactory from the standpoint of loading by the gates and providing sufficient current in the load.

Voltage and current imbalance may result from one or more of the following conditions: 1. variations in flip-flop or buffet output voltages, 2. nonuniformity of resistors due to their tolerances, 3. finite voltage drop across conducting diodes, 4. diode leakage currents or presence of bias resulting in partial conduction, and 5. loading by the detector which senses the output current. In general, the effect of these imbalances is a reduction in magnitude of the zero and 1 current outputs. A few typical examples will be discussed which illustrate the practical problems involved in designing current-operated diode logic circuits.

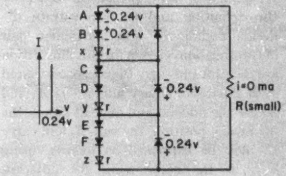
A simple 2-input add gate, shown in Fig. 2(B), serves to illustrate the effect of unequal current sources. Assume the inputs as indicated on the diagram, but let the magnitude of source A be 0.4 ma, and determine the remaining currents. The diode opposite source A will remain nonconducting, but the output current will be reduced to 0.4 ma because source A is in series with the load. Current source B' will deliver its specified 0.5 ma, of which 0.4 ma must flow through the load and 0.1 ma must flow through the diode faced by the source. This diode will conduct, although theoretically it is to remain nonconducting. For this simple network the effect of the unbalanced current is almost trivial; it simply reduces the output current. In more elaborate networks the conducting diodes would establish "sneak paths" which could adversely affect the performance of the circuit. No reason was given for the inequality in the assumption made earlier, that the currents were unequal. When the inequality was due to unequal input voltages or currents, the previous arguments would apply only if the errors were symmetrical. This brings up the problem of errors due to voltage imbalance, which is perhaps the more serious problem because it applies to all types of networks.

If the currents are derived as was suggested, with flip-flops and series resistances, all current sources will essentially be commonly connected. Assume the flip-flop outputs to be +E and −E, symmetrical about zero volts or ground, connect identical resistances to each output and then connect these resistances to a load. If the load were a short circuit, the potential of this point with respect to ground would be zero volts. Hence, if the inputs are symmetrical with respect to ground it would be desirable to have all portions of the network at this potential. This requirement cannot be satisfied for all input states of a network due to the finite voltage drop exhibited by a conducting, nonideal diode.

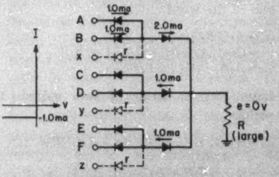
To emphasize this condition, consider a 20-input multiply gate of which 19 inputs are in state 1. The cumulative voltage across the 19 conducting diodes would be approximately 4.5 volts. It is obvious that this would severely unbalance the currents if the input voltage were ±5.0 volts. Furthermore, the terminal voltage of the current source associated with the zero input would have to exceed this voltage (and be of opposite polarity) to cause the current in the load to flow in the proper direction. A solution to this problem, using a multistage network, consists of dividing the network into several multiply gates each having only four or five inputs. The second stage would also be a multiply gate, and its inputs would be the outputs of the first-stage gates. In terms of circuit particulars, equal groups of diodes would be shunted by a single diode, pointed in the same direction. This procedure could be extended to several stages if necessary.

Multistage networks are generally formed by cascading alternately add and multiply gates. The circuit will therefore contain several columns of diodes, those in adjacent columns oppositely polarized. Over-all voltage imbalance is usually less severe in the typical case because the voltage drops across the diodes pointed up will tend to be canceled by the drops across diodes pointed down. It has been demonstrated that practical multistage networks can be designed by means of current-operated techniques.

Factored transmission expressions require fewer literals than the equivalent lower-order functions. Since each literal represents a required current source input, higher-order functions require fewer inputs. Any reduction in the number of inputs is advantageous in limiting the detrimental effects of voltage and current imbalance. All items discussed thus far may be classified as engineering prob-



(A)CURRENT-OPERATED GATE



(B) VOLTAGE-OPERATED GATE

Fig. 14. Realization requiring redundant diodes

lems. In most cases several acceptable solutions exist and consideration must be given to the specific application or requirement.

A second category of problems requires modification of the realization rules because the resulting networks fail to meet the transmission specifications. In particular, the output currents may be of the wrong polarity or greatly reduced in magnitude. Why do the realization rules fail? Previous discussion has indicated that the auxiliary series bias-voltage sources shown in Fig. 3(B) could be eliminated because the voltage drop across conducting diodes is of sufficient magnitude to reverse bias diodes which are to remain nonconducting. The elimination of the bias-voltage sources is usually valid and the practical advantages of a technique which employs diodes only but does not require numerous floating voltage sources is obvious. Although the elimination of the bias-voltage sources is valid for all single-stage gates it is not valid for all multistage gates.

To review the requirements of multistage gates, it will be recalled that the output of all stages must effectively represent current sources (for every possible state of the inputs) to drive the succeeding stages properly. The output of a stage will fail to represent a current source if for any specified input state all diodes in the stage become conducting. The current sources forward-bias the diodes in an attempt to establish sufficient terminal voltage (of the proper polarity) to cancel opposing voltage drops present in the mesh of interest. This voltage must be supplied by the current sources in order to establish current flow in the correct direction. Current-operated diode logic gates exhibiting the conditions outlined above may not satisfy the desired transmission specifications.

However, experience gained in designing the static multiplier (mentioned at the start of the paper), and numerous other gates, indicates such difficulties to be the exceptions and not the rule. Furthermore, if only minor network modifications are required to correct the transmission, it is simpler to treat malfunctioning realizations as special cases. The remedy usually requires the addition of one or more diodes strategically located. These additional diodes will be referred to as redundant diodes because they perform no "logical" function. In certain instances, the addition of a redundant diode permits, oddly enough, the omission of another diode which appears to be "logically" required. Diodes which may be omitted without affecting the transmission of the gate will be referred to as superfluous diodes.

The first class of networks to be discussed is that of 2-stage realizations representative of the minimum-sum or minimum-product forms of the transmission function. Fig. 14(A) shows the current-operated realization of the function $T = AB + CD + EF$; a sketch of the $V-I$ characteristics representing the equivalent circuit of a nonideal diode is included for reference. The dual of this network is found in Fig. 14(B) which includes a sketch of the V-I characteristics of the equivalent "dual diode." This hypothetical element conducts for all values of positive applied voltage and exhibits a constant "leakage current" for negative voltage. An equivalent circuit of this device consists of an ideal diode in parallel with a constant current source, polarized in the direction to cause the diode to conduct. If a practical device having these characteristics were available, voltage-operated gates could be realized in a manner dual to that suggested for current-operated gates.

The reason for introducing this hypothetical diode and the corresponding network is to provide an alternate approach to clarify the following discussions. The networks are identical to the circuits of voltage-operated gates with all stage load resistors omitted. Familiarity with voltage-operated gate circuits greatly simplifies the problem of determining the proper location of redundant diodes.

Return to Fig. 14(A) and assume the input state $A = B = 1$ and $C = D = E = F = 0$. The output should be 1, because of current down through the load. The measured output current of this circuit will be approximately zero for this particular input state, in accordance with the idealized calculations. For current to flow down through the load, the cumulative terminal voltage of current sources A and B must exceed 0.48 volt to overcome the voltage drop of the two conducting diodes in the second stage. However, the diodes faced by sources A and B will conduct as soon as the terminal voltage equals 0.24 volt, and will limit the terminal voltage to this value. Therefore, the voltage drops across the conducting diodes will exactly cancel and no current will flow in the load. The identical results will be obtained if only C and D or E and F are equal to 1.

Several solutions suggest themselves; one is to use six silicon diodes in the first stage and three germanium diodes in the second stage. A second method requires the addition of three redundant diodes, one in each first-stage gate. These diodes are shown by dashed lines in Fig. 14(A); the r indicates that they are redundant. The associated sources are labeled x, y, and z for identification, to facilitate discussing the various ways in which they may be used. The diodes effectively simulate the bias-voltage sources which were previously eliminated.

To complete this simulation, sources x, y, and z could simply be unit current sources causing the associated diodes to conduct, exhibiting a voltage drop of 0.24 volt. Other alternatives exist: source x could be either a second source A or B, source y could be either a second source C or D, and source z could be either a second source E or F. Any of the suggested solutions are satisfactory; the decision is left to the designer.

For those more familiar with voltage-operated gates a brief discussion of Fig. 14(B) will demonstrate the utility of the analogy. Assume the same input state as before. For ease of calculation let the current sources shunting the ideal diodes be equal to 1.0 ma, and determine the output current. The diodes connected to voltage sources A and B are both reverse biased. A current of 1.0 ma will flow bypassing the diodes (due to the current sources) with the result that 2.0 ma will enter the next stage. The 1.0 ma associated with the diode in the second stage is confined to circulate in a mesh which consists of this diode and its corresponding source and does not enter into the calculations.

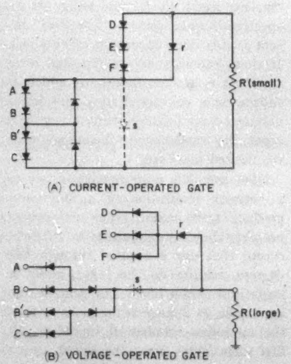
Voltage sources C, D, E, and F hold their respective inputs low, and allow the diodes to conduct. However, the 2.0 ma leaves the top diode in the second stage, and divides as shown. Each of the 1.0-ma currents again divides and enters its respective voltage source. The load current is zero and no output is obtained. If these calculations are repeated after the three redundant diodes are added as shown, each connected to a unit input voltage source or duplicating an existing input, the transmission will conform with the desired function.

Obviously, the same remarks apply to a network realization of a minimum-product transmission. The uniqueness property shows the circuits to be identical except that one is rotated through 180 degrees, relative to the other, about a horizontal center line. The general rule applicable to 2-stage realizations is that the minimum number of inputs to all first-stage gates must be equal to or greater than the number of inputs to the second-stage gates. From a geometrical standpoint this implies that there must be at least as many series-connected diodes in each gate in the first column as there are diodes in the second column.

The second class of networks includes all multistage realizations which do not satisfy the desired transmission specifications because the series bias-voltage sources were eliminated. This miscellaneous category also includes networks which may be further minimized, not by algebraic methods but from operational considerations which show that a particular diode is not required to conduct for any state of the inputs. The network in Fig. 15(A) and its dual in Fig. 15(B) will be used to illustrate these points. These circuits each contain a diode which is superfluous, and is shown by dashed line and designated by an *s*. In the current-operated gate this diode may be removed and replaced by an open circuit, and in the voltage-operated gate it may be removed and replaced by a short circuit.

Experimentation showed this gate to malfunction for the input state $A=B=C=0$ and $D=E=F=1$. The desired transmission function is $T=DEF\,(AB+B'C)$ and for the given input state the transmission should be equal to zero but was found to be 1. The reason for the malfunction is that insufficient voltage was established by the first stage sources, due to the presence of the two series diodes in the second stage, to overcome the cumulative voltage developed across the conducting diodes faced by generators *D*, *E*, and *F*.

Again, many acceptable solutions exist and the designer must be cautioned against overlooking the simple approach. A straightforward solution is obtained by expanding the function and then synthesizing the result, $T=DEFAB+DEFB'C$. This requires two more diodes and three additional inputs. The solution shown in the diagram was selected because the network is more minimal and it illustrates the presence of a superfluous diode. Adding the redundant diode solves the problem because it reduces, by a factor of three, the voltage across the three input diodes faced by sources *D*, *E*, and *F* when all inputs are in state 1.



(A) CURRENT-OPERATED GATE

(B) VOLTAGE-OPERATED GATE

Fig. 15. Realization requiring redundant elements and detection of superfluous diodes

The superfluous diode, "logically" required by the realization rules, may be eliminated since all four input sources *A*, *B*, *B′*, and *C* cannot simultaneously be equal to 1 requiring the presence of the path provided by this diode. Currents flowing between the nodes to which the superfluous diode is connected are required to choose paths established in the network to the left of this diode because of the disjunctive nature of the function. The net result of this solution requires no more diodes but corrects the malfunction. Analysis of the voltage-operated gate parallels that of the previous discussion and is left to the reader.

One last remark concerning this realization must be made because it occurs frequently. If a hybrid realization were considered, the portion of the function represented by $AB+B'C$ could be synthesized with the use of a lattice consisting of four diodes. This in addition to the four diodes required to form the term *DEF* would complete the circuit. The hybrid circuit was constructed and the transmission experimentally verified. It satisfied the desired transmission equation with a total of eight diodes. Experience has proved, in many instances, that the more minimal network operates satisfactorily, whereas a less minimal realization may exhibit malfunctions for certain input states.

Construction of a Static Multiplier

A static multiplier is a device capable of simultaneously multiplying two numbers without recourse to the accumulation of partial products or shifting operations. Signals representing the multiplicand and the multiplier are simultaneously applied to the logic gate inputs and, after the transients in the system have disappeared, signals representing the product appear at the outputs. It was decided that the design of a static multiplier would provide ample opportunities to evaluate both the design techniques and the advantages of current-operated gates. To achieve this goal, a device capable of multiplying two 4-bit binary numbers was designed and built.

				A	B	C	D
				a	b	c	d
				Ad	Bd	Cd	Dd
			Ac	Bc	Cc	Dc	
		Ab	Bb	Cb	Db		
	Aa	Ba	Ca	Da			
K_{64}	K_{54}	K_{44}	K_{34}				
K_{72}	K_{62}	K_{52}	K_{42}	K_{32}	K_{22}		
S_8	S_7	S_6	S_5	S_4	S_3	S_2	S_1

Fig. 16. General procedure for binary multiplication

The operation of the multiplier is conceptually simple because it parallels the familiar decimal system method. Fig. 16 illustrates the general procedure, and Appendix II lists the pertinent carry and sum functions. The first form of the expressions is derived by inspection, using Fig. 16, and the second form is the result after minimizing, and represents the transmission function actually realized. In order to limit the number of current buffers required, the minor products were not separately formed with the exception of *Dd*, which is the first sum digit.

Two auxiliary functions were incorporated to reduce the complexity of three gate networks; the functions involved are K_{42}, K_{44}, and S_4. Several redundant superfluous diodes were incorporated, in the networks which represented the realization of K_{42}, K_{44}, and K_{02}, which were not specified by straightforward application of the realization rules. However, the modified circuits performed satisfactorily. Approximately 270 diodes were required to construct the 20 gates used in the static multiplier. A total of four redundant diodes were included to correct malfunctions exhibited in special cases. To balance this number exactly, it was found that four diodes could be omitted because they were superfluous. The previous statements are submitted as evidence that malfunctioning realizations are infrequently encountered and that the network modifications which may be required are usually minor.

The static multiplier block diagram of Fig. 17 indicates the flow of information and depicts the major areas included in the system. Static tests performed on the equipment indicated that all gates operated satisfactorily for the 256 different inputs, thus verifying the correctness of the functions, the minimization steps,

Vol.
Pt. 1
1960

the realization rules, the actual circuits, and the wiring. Dynamic tests were made with a single flip-flop operating in the commutative mode to obtain data on carry propagation time to ascertain the maximum operating speed of the static multiplier.

As always, it is a difficult matter to specify the "worst case" situation. It was found that the significant delays could be attributed to the response time of the current buffer circuit. Therefore, the propagation times were measured for those input combinations which required the greatest number of current buffers to change state. The longest delay measured was 2.4 μsec (microsecond) for a combination which required 12 current buffers to change state. Using the 0.2-μsec delay figure, determined from tests of the current buffer, and multiplying this by 12 because of the number of buffers which made a transition, a total delay of 2.4 μsec is obtained. This value is in agreement with the experimental results. From these data it may be deduced that the maximum reliable operating speed of the static multiplier is approximately 3.0 μsec, which corresponds to 333,000 multiplications per second.

Conclusions

Two realization methods have been presented which are applicable to the design of current-operated diode logic gates. The dual method, as the name implies, consists of a set of rules which are analogous to the procedures employed for voltage-operated circuits. In single-stage gates obtained by using the dual method the input current sources are connected in shunt with their respective diodes, and the diodes are connected in series to constitute the network. The state of a variable is represented by the direction of the current. The output of the gate is also a current which may be used as an input to a succeeding state.

Multistage networks may be realized by series-parallel arrays of basic gate circuits. The lattice method, which has no voltage-operated counterpart, is based on the transmission characteristics of a lattice network of four series-connected diodes. The general form of the transmission function realized by the lattice is $T=T_aT_b+T'_aT_c$. It is observed that the operation of T_a is one of selection. For the special situation where $T_c=T'_b$, the lattice performs the operation of sum modulo 2. Multistage networks may be obtained by cascading a number of lattice circuits. Hybrid realizations which combine the advantages of both procedures, generally require fewer diodes than networks obtained by either method alone. Discriminating application of the current-operated realization rules, which offer numerous synthesis alternatives, should yield more minimal networks than heretofore possible using voltage-operated techniques.

It has been shown that the realization of a transmission function is identical to the realization of the complement of that function by current-operated methods. This unique property exists because the direction of the output current is functionally related to the direction of the input currents.

Although current-operated techniques are not the "last word" in the realization of switching functions using diode networks, this approach does suggest some new and interesting possibilities.

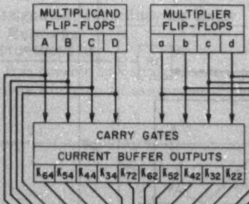


Fig. 17. Static multiplier block diagram

Appendix I. Conventions and Design Procedures

Conventions

1. The literal adjacent to the diode denotes the variable associated with the implicit current source impressed across the diode. When the state of the input represented by the literal is 1, the current source points up. Conversely, when the state of the input is zero, the current source points down.

2. The transmission is 1 when the direction of current flow through the load is down; conversely, a zero is obtained when the current through the load is directed up.

Dual Method

Basic Single-Stage Gates

1. Series-connected diodes pointing up, in the column adjacent to the literals which identify the current sources, represent the elements of an add gate.

2. Series-connected diodes pointing down, in the column adjacent to the literals which identify the current sources, represent the elements of a multiply gate.

Multistage Gates

1. Series-connected diodes pointing up represent the elements of add gates. The current input to each diode is either the output of a previous stage or the output of a primary current source.

2. Series-connected diodes pointing down represent the elements of multiply gates. The current input to each diode is either the output of a previous stage or the output of a primary current source.

Lattice Method

Basic Gates

1. The transmission of the basic lattice network is represented by the selection function, $T=AB+A'C$.

2. In accordance with the conventions of Fig. 7(B), the transmission is equal to the product of the central literal and the lower literal added to the product of the complement of the central literal and the upper literal.

Multistage Gates

1. The transmission of the general lattice network is represented by the selection function $T=T_aT_b+T_a'T_c$.

2. The current inputs are either the output of a previous stage or the output of a primary current source.

Appendix II. Static Multiplier Transmission Expressions

Carry Functions

$K_{22}=CDcd$

$K_{32}=K_{34}'[K_{22}(Db+Cc+Bd)+Db(Cc+Bd)+CcBd]$
$\quad =K_{34}'K_{22}+K_{34}'(CDbc+BDbd+BCcd)$

$K_{42}=K_{44}'[K_{32}(Da+Cb+Bc+Ad)+Da(Cb+Bc+Ad)+Cb(Bc+Ad)+BcAd]$
$\quad =K_{44}'[K_{01}(K_{32}\oplus S_4)'+K_{32}(K_{32}\oplus S_4)+CDab+ABcd]$

$K_{52}=K_{54}'[K_{42}(K_{34}+Ca+Bb+Ac)+K_{32}(Ca+Bb+Ac)+Ca(Bb+Ac)+BbAc]$
$\quad =K_{54}'[Ca(K_{42}+Bb)+Bb(K_{42}+Ac)+Ac(K_{42}+Ca)]$

$K_{62}=K_{64}'[K_{52}(K_{44}+Ba+Ab)+K_{44}(Bb+Ab)+BaAb]$
$\quad =K_{64}'K_{52}(K_{44}+Ba+Ab)+K_{64}'ABab$

$K_{72}=K_{62}(K_{64}+Aa)+K_{64}Aa=K_{62}Aa$

$K_{34}=K_{22}DbCcBd=K_{22}Bb$

$K_{44}=K_{32}[Da(CbBc+CbAd+BcAd)+CbBcAd]+DaCbBcAd$
$\quad =K_{02}+BCbc(Dad'+AD'd)$

$K_{54}=K_{42}[K_{34}(CaBb+CaAc+BbAc)+CaBbAc]=K_{34}(A+a)+K_{42}ABCabc$

$K_{64}=K_{52}K_{44}ABab$

Sum Functions

$S_1=Dd$

$S_2=Cd\oplus Dc$

$S_3=Bd\oplus Cc\oplus Db\oplus K_{22}$

$S_4=Ad\oplus Bc\oplus Cb\oplus Da\oplus K_{34}$
$\quad =K_{01}\oplus Cb\oplus Da\oplus K_{34}$

$S_5=Ac\oplus Bb\oplus Ca\oplus K_{34}\oplus K_{42}$

$S_6 = Ab \oplus Ba \oplus K_{44} \oplus K_{52}$
$S_7 = Aa \oplus K_{54} \oplus K_{62}$
$S_8 = K_{64} \oplus K_{72}$

Auxiliary Functions

$K_{01} = Ad \oplus Bc$
$K_{02} = ADad[BC(b+c)+bc(B+C)]$

References

1. SWITCHING CIRCUITS AND LOGICAL DESIGN (book), S. H. Caldwell. John Wiley & Sons, Inc., New York, N. Y., 1958.

2. INTRODUCTORY CIRCUIT THEORY (book), E. A. Guillemin. John Wiley & Sons, Inc., 1953.

3. SYNTHESIS OF PASSIVE NETWORKS (book). E. A. Guillemin. John Wiley & Sons, Inc., 1957.

4. THE MAP METHOD FOR SYNTHESIS OF COMBINATIONAL LOGIC CIRCUITS, M. Karnaugh. *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 72, Nov. 1953, pp. 593–99.

5. A FUNCTIONAL CANONICAL FORM, H. A. Curtis. *Journal*, Association for Computing Machinery, Baltimore, Md., vol. 6 no. 2, Apr. 1959, pp. 245–58.

Assignment of Carry Variables in Iterative Networks

E. J. McCLUSKEY, JR.
NONMEMBER AIEE

BOTH SEQUENTIAL circuits (Fig. 1) and iterative circuits (Figs. 2 and 3) are usually designed by first forming a stable table (flow table) describing the desired circuit performance.[1–3] In this table the various combinations of signals on the feedback or carry leads (corresponding to the internal states of the circuits) are represented by decimal numbers. It is necessary to assign specific combinations of binary signals to these (decimal) states before the actual circuit design can be carried out. This corresponds to the transition from a state table, such as Table I—A to an output table, such as Table I—B. For a given table, different assignments will usually lead to different circuits which can vary significantly in cost. The problem of minimizing the cost of the equipment can be even more important for iterative circuits since each "cell" occurs many times in the final circuit. The number of different assignments is so large that it is not feasible to try all different assignments for circuits having more than two internal variables.[4] This particular problem of discovering assignments which correspond to economical circuits is still unsolved.

For certain types of iterative or sequential circuits there are restrictions on the assignments since certain assignments lead to circuits which are either not physically realizable or which do not perform according to the original state table. For example, an asynchronous sequential circuit must have an assignment in which either no more than one feedback loop is unstable at one time, or, if several loops are unstable simultaneously, the final stable state reached must be independent of the order in which the feedback loops change their conditions.[1,2]

In an iterative network constructed of contacts it is not true that any assignments will lead to a physically realizable network. For example, the function $Y_1 = y_1 y_2$ cannot be realized by a contact network. The symbols y_1 and y_2 represent inputs to a typical cell and will be called *input-carry variables* since they correspond to signals on the "carry" leads between cells. The symbol Y_1 represents an output from a typical cell and will be called an *output-carry variable*. The circuit inputs will be denoted by x or x_1, x_2 ..., and the circuit outputs by Z, or Z_1 Z_2, ... A specific combination of signals on the input-carry leads will be called an *input-carry state*, denoted by s_j; and a specific combination of signals on the output carry leads will be called an *output-carry state* and denoted by S_j.

The reason that $Y_1 = y_1 y_2$ cannot be realized by a contact network is that y_1 and y_2 represent signals on two separate leads; with the use of contacts only it is impossible to produce a signal on a third lead which is present only when signals are present on both the y_1 and y_2 leads. For similar reasons, it is impossible to realize a function such as $Y_1 = xy_1y_2$. The other type of function which cannot be realized is one which contains the complement of an input-carry variable: $Y_1 = x(y_1 + y_2')$. This is because it is impossible, with contacts only, to produce a signal on one lead whenever a signal is not present on another lead. In general, the following rule applies:

> Any function which contains the product of two or more y_i's or contains any y_i', cannot be realized in an iterative contact network.

A carry assignment for an iterative contact network must result in output and output-carry functions which satisfy this requirement.

It is possible to have a set of output-carry functions which satisfy the conditions required for a contact network realization and which are still not physically realizable. For example, the typical cell of Fig. 4 is derived from the output

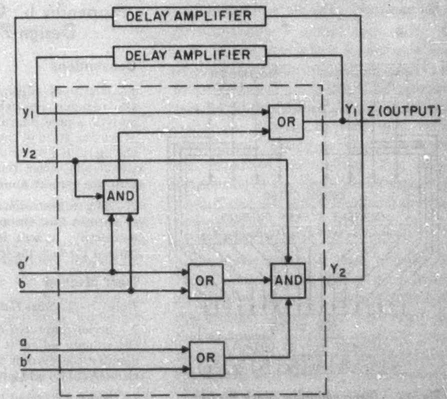


Fig. 1. Sequential circuit which has an output only if A<B:

$A = a_n 2^n + a_{n-1} 2^{n-1} + \ldots a_1 2^1 + a_0 2^0$
$B = b_n 2^n + b_{n-1} 2^{n-1} + \ldots b_1 2^1 + b_0 2^0$

Paper 60-1211, recommended by the AIEE Computing Devices Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Chicago, Ill., October 9–14, 1960. Manuscript submitted May 27, 1960; made available for printing August 5, 1960.

E. J. McCLUSKEY, JR., is with Princeton University, Princeton, N. J.

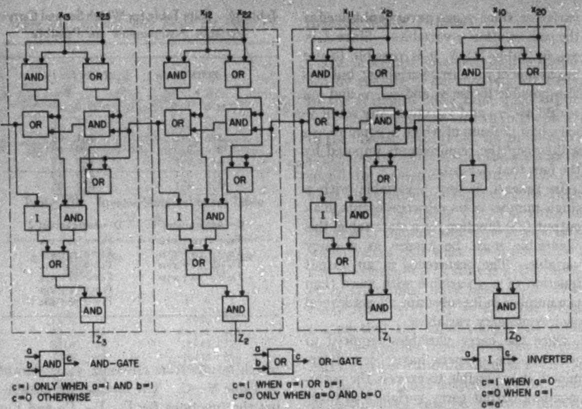


Fig. 2. Iterative network for adding two binary numbers X_1 and X_2:

$$X_1=2^0x_{10}+2^1x_{11}+2^2x_{12}+2^3x_{13} \quad Z=X_1+X_2$$
$$X_2=2^0x_{20}+2^1x_{21}+2^2x_{22}+2^3x_{23}$$

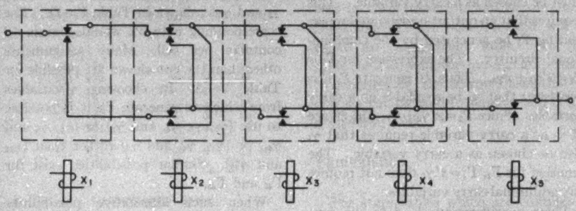


Fig. 3. Relay iterative network, closed when an odd number of relays are operated

table of Table II—B. Since this typical cell contains a "sneak path" ($x'y_1$) in Y_3, a network constructed from it will not function properly. The sneak path cannot be avoided (with the use of contacts only) and a different assignment of carry variables must be used in order to obtain a correct network.

Assignment Procedure for Relay Networks

It is possible by a straightforward procedure to obtain an assignment of carry variables appropriate for a contact network. The resulting assignment will specify output-carry and output functions which do not contain complements or products of input-carry variables. The assignment procedure itself does not guarantee the absence of sneak paths but it is possible to rule out such situations, by a simple modification of the basic procedure.

To facilitate the discussion, the following notation will be used:

γ_i=a variable equal to one only when the input-carry state is i.

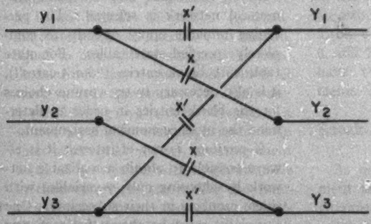


Fig. 4. Typical cell of iterative contact network with a sneak path

Table I. Tables for Fig. 1

A—Flow Table

Initial state		ab 00	01	11	10	z (output)
$A=B$	1	1	3	1	2	0
$A>B$	2	2	2	2	2	0
$A<B$	3	3	3	3	3	1

B—Output Table

Initial state y_1y_2	ab 00	01	11	10	z (output)
01	01	10	01	00	0
00	00	00	00	00	0
10	10	10	10	10	1
	Y_1Y_2				

$Y_1=y_1+y_2a'b;\ Y_2=y_2(a+b')(a'+b);\ z=Y_1$

γ_{ijk}=a variable equal to one only when the input-carry state is i or k or j, etc.

Γ_i=a variable equal to one only when the output-carry state is i, etc.

Thus, specifying $y_1=\gamma_2$, $Y_1=\Gamma_2$, $y_2=\gamma_{13}$, $Y_2=\Gamma_{13}$, $y_3=\gamma_{235}$ and $Y_3=\Gamma_{235}$ is equivalent to the following table:

s_k S_k	y_1 Y_1	y_2 Y_2	y_3 Y_3
1	0	1	0
2	1	0	1
3	0	1	1
4	0	0	0
5	0	0	1

If two γ-variables are added, the resulting sum variable has for its subscript all numbers which appear in the subscripts of either (or both) of the γ-variables which were added. Thus, $\gamma_{12}+\gamma_{35}=\gamma_{1235}$, $\gamma_{12}+\gamma_{23}=\gamma_{123}$. This is true since the sum of a variable which equals one only for row i (γ_i) and a variable which equals one for row j (γ_j) is a variable which equals one for both row i and row j (γ_{ij}), etc.

In terms of the γ-variables, the assignment of carry variables consists of selecting a set of γ-variables to be used as the carry variables. In the actual iterative network, signals which correspond to the selected γ-variables will be present on the carry leads. In general, signals corresponding to other γ-variables will be produced by the combinational circuitry.

In a contact network there is a restriction on the signals produced by the combinational circuitry: they can correspond only

Table II. Tables for Fig. 4

A—State Table

s_j	x 0	1	Z
1	1	2	0
2	5	3	0
3	4	5	1
4	4	2	1
5	5	5	0
	S_j		

B—Output Table

$y_1y_2y_3$	x 0	1	Z
1 0 0	1 0 0	0 1 0	0
0 1 0	0 0 0	0 0 1	0
0 0 1	1 0 1	0 0 0	1
1 0 1	1 0 1	0 1 0	1
0 0 0	0 0 0	0 0 0	0
	$Y_1Y_2Y_3$		

$Y_1=x'(y_1+y_3),\ Y_2=xy_1,\ Y_3=x'y_3+xy_2,\ Z=y_3$

Table III. Illustrative State Tables

A—Circuit Output Table

	x_i			
s_j	0	1	Z_1	Z_2
1	1	2	0	1
2	2	3	0	1
3	3	4	0	1
4	4	5	0	1
5	5	5	1	0

S_j

B—Cell Output Table

	x_i	
s_j	0	1
1	1, 01	2, 01
2	2, 01	3, 01
3	3, 01	4, 00
4	4, 00	4, 10

S_j, Z_1Z_2

Table IV. Illustrative State Table

	x		
s_j	0	1	Z
1	1	2	0
2	2	3	0
3	3	4	1
4	4	4	0

S_j

to γ-variables which can be obtained by adding together some of the γ-variables used as carry variables.

This is true since it is impossible to complement or multiply carry variables in a contact network. For example, if the assignment $y_1=\gamma_2$, $y_2=\gamma_{13}$, and $y_3=\gamma_{235}$ is used, a contact network would generate signals corresponding to $\gamma_{123}=\gamma_2+\gamma_{13}$, and $\gamma_{1235}=\gamma_{13}+\gamma_{235}$, but not to γ_3 or γ_4 or γ_{23}, etc. The following discussion will apply only to iterative contact networks.

It is possible, without assigning carry variables, to write an expression for the outputs (Z_i) in terms of γ-variables. For Table III—A: $Z_1=\gamma_5$ and $Z_2=\gamma_{123}$; and for Table III—B: $Z_1=x\gamma_4$ and $Z_2=x'\gamma_{123}+x\gamma_{12}=\gamma_{12}+x'\gamma_3$. A set of γ-variables to be used as carry variables must be chosen so that the following obtains:

> All γ-variables which appear in the expressions for the outputs (Z_i) are either carry variables or are equal to sums of carry variables.

This is necessary since it is possible to generate only signals which correspond to carry variables or to sums of carry variables. In Table III—A, $Z_1=\gamma_5$, so that γ_5 must be chosen as one of the carry variables, since γ_5 can never be obtained as the sum of other γ-variables. Since $Z_2=\gamma_{123}$ in Table III-A, it is possible to use γ_{123} as one of the carry variables; but it is also possible to use γ_1 and γ_{23}, γ_2 and γ_{13}, γ_3 and γ_{12}, γ_{13} and γ_{23}, etc. Any set of γ-variables the sum of which is equal to γ_{123} will satisfy the requirements imposed by the fact that $Z_2=\gamma_{123}$.

In general, if any γ-variable with a single number in its subscript occurs in an output (Z_i) function for a state table, this γ-variable must be chosen as a carry variable. The occurrence in an output function of a γ-variable with more than one number in its subscript allows several choices of carry variables.

Once the carry variables required to generate the outputs have been determined, it is possible to express the corresponding output carry variables as functions of γ-variables, and from these expressions to determine which additional carry variables are necessary, etc. For example, in Table IV, $Z=\gamma_3$ so that γ_3 must be chosen as a carry variable. This requires that an output-carry signal representing Γ_3 be generated by the combinational circuitry. The expression for Γ_3 is $\Gamma_3=x'\gamma_3+x\gamma_2$. Thus, to generate Γ_3 it is necessary that γ_2 be available as a carry variable. Since $\Gamma_2=x'\gamma_2+x\gamma_1$, the choice of γ_2 as a carry variable requires that γ_1 also be chosen as a carry variable. The function for Γ_1, $\Gamma_1=x'\gamma_1$ does not require any additional carry variables.

These considerations lead to the conclusion that the choice of γ_3, γ_2, and γ_1 as carry variables is an appropriate assignment for Table IV. Moreover, this is the only possible assignment since no alternative possibilities occurred in the process of selecting the carry variables. A typical cell for the final network is shown in Fig. 5.

For the state table of Table V—A several different assignments are possible and the amount of equipment in the final network varies for different assignments. For this table, $Z=\gamma_{56}$ so that either γ_{56} or γ_5 and γ_6 must be chosen as carry variables. The carry-variable assignment which results from choosing γ_{56} is shown in Table V—C. The method for obtaining the functions for Γ_{56}, Γ_{12}, Γ_{45} and Γ_{23} is illustrated in Table V—B.

Table V. State Table for Which Several Carry-Variable Assignments are Possible

A—State Table

	x		
s_j	0	1	Z
1	6	1	0
2	5	2	0
3	4	3	0
4	1	6	0
5	2	5	1
6	3	4	1

S_j

Table for Γ_{56}

	x	
s_j	0	1
1	1	0
2	1	0
3	0	0
4	0	1
5	0	1
6	0	0

C—Assignment 1

$\Gamma_{56}=x'\gamma_{12}+x\gamma_{45}$
$\Gamma_{12}=x'\gamma_{45}+x\gamma_{12}$
$\Gamma_{45}=x'\gamma_{23}+x\gamma_{56}$
$\Gamma_{23}=x'\gamma_{56}+x\gamma_{23}$

D—Assignment 2

$\Gamma_6=x'\gamma_1+x\gamma_4$
$\Gamma_5=x'\gamma_2+x\gamma_5$
$\Gamma_1=x'\gamma_4+x\gamma_1$
$\Gamma_4=x'\gamma_3+x\gamma_6$
$\Gamma_2=x'\gamma_5+x\gamma_2$
$\Gamma_3=x'\gamma_6+x\gamma_3$

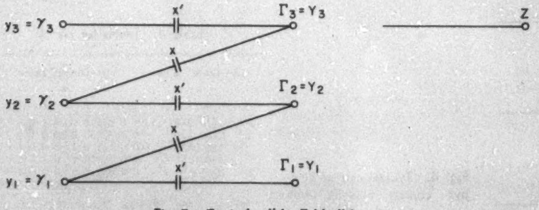


Fig. 5. Typical cell for Table IV

This assignment leads to a network in which there are eight contacts per cell. The assignment which results from using γ_5 and γ_6 is shown in Table V—D. The corresponding network would contain 12 contacts per cell. Many assignments other than the two shown are possible for Table V—A. In choosing γ-variables from which to generate Γ_{56} it is possible to use (γ_1, γ_2, γ_4, and γ_5) or (γ_1, γ_2 and γ_{45}) or (γ_{12}, γ_4, and γ_5) rather than (γ_{12} and γ_{45}). Similar possibilities exist for Γ_{45} and Γ_{23}.

When such alternative possibilities occur for choosing γ-variables, a suitable assignment (except possibly for sneak paths) for an iterative contact network will result from each of the different possibilities. Thus the procedure just described always results in a suitable assignment.

There is no technique known for determining which of several possible choices of γ-variables will result in the most economical network. If economy is important, it is necessary to carry out the procedure several times until all the various alternatives have been considered. For each assignment, the amount of equipment required in the corresponding network is determined and the most economical network is selected. The preceding remarks apply directly to completely specified state tables. For state table with blank entries ("don't-cares"), it is also necessary to try various choices for the blank entries in order to determine the most economical assignment.

If economy is not of interest it is always possible to obtain a realizable network by choosing only γ-variables with single numbers in their subscripts. One γ-variable must be chosen for each row

Table VI. Inverse State Table for Table IV

	x	
S_j	0	1
1	1	–
2	2	1
3	3	2
4	4	3, 4
Z	3	3

S_j

Table VII. Inverse State Table for Table V—A

A—Inverse State Table

	x	
S_j	0	1
1	4	1
2	5	2
3	6	3
4	3	6
5	2	5
6	1	4
Z	56	56

s_j

B—Assignment Table

	x	
Γ	0	1
Z	56	56
56	12	45
12	45	12
45	23	56
23	56	23

γ

Table VIII. Equivalence Determined by Assignment Procedure

A—State Table

	x		
s_j	0	1	Z
1	3	7	0
2	4	8	0
3	5	1	0
4	6	2	0
5	7	5	1
6	8	6	1
7	1	3	1
8	2	4	1

S_j

B—Inverse State Table

	x	
S_j	0	1
1	7	3
2	8	4
3	1	7
4	2	8
5	3	5
6	4	6
7	5	1
8	6	2
Z	5678	5678

s_j

C—Assignment Table

	x	
	0	1
Z	5678	5678
5678	3456	1256
3456	1234	5678
1256	3478	3456
1234	1278	3478
3475	1256	1278
1278	5678	1234

D—Carry Variables

S_j	γ_{5678}	γ_{3456}	γ_{1256}	γ_{1234}	γ_{3478}	γ_{1278}
1	0	0	1	1	0	1
2	0	0	1	1	0	1
3	0	1	0	1	1	0
4	0	1	0	1	1	0
5	1	1	1	0	0	0
6	1	1	1	0	0	0
7	1	0	0	0	1	1
8	1	0	0	0	1	1

of the state table, although for some tables one of these γ-variables may prove to be not actually necessary. (A terminal state for which there are no outputs equal to one can be assigned the combination of carry variables for which all carry variables are equal to 0). Since any γ-variable (e.g., γ_{13}) can be obtained as a sum of γ-variables with single numbers in their subscripts ($\gamma_1+\gamma_3$), an assignment using all γ-variables with single numbers in their subscripts can realize the functions required for any state table. This is the so-called one-hot coding.

Inverse State Tables

The procedure just described is simplified by the use of an inverse state table. In the standard state table the output-carry states from the entries in the table. In the inverse state table, the input-carry states are the entries, and the rows correspond to output-carry states. The inverse state table corresponding to Table IV is shown in Table VI. The entry in the ($x=1$) column and ($S_j=2$) row is 1, which means that when the input is $x=1$ and the output-carry state is 2, the input-carry state must have been $s_j=1$. The entry for $x=1$, $S_j=4$ is 3,4, which means that the input-carry state could have been either 3 or 4 when the output-carry state is 4 and the input is $x=1$.

The output and output-carry functions of γ-variables can be written directly from the table. Thus $\Gamma_1=x'\gamma_1$, $\Gamma_2=x'\gamma_2+x\gamma_1$, $\Gamma_3=x'\gamma_3+x\gamma_2$, $\Gamma_4=x'\gamma_4+x\gamma_{34}$, $Z=\gamma_3$. By combining two or more rows, the Γ-functions for Γ's with several numbers in their subscripts can be obtained. See Table VII—B. Such a table will be called an "assignment table." In this table the Z-row shows that γ_{56} can be used to generate Z; the 56-row shows that γ_{12} and γ_{45} can be used to generate Γ_{56} ($\Gamma_{56}=x'\gamma_{12}+x\gamma_{45}$), etc.

Equivalence

For state tables which are completely specified (containing no don't-cares), it is unnecessary to test for equivalent states with this procedure for assigning carry variables. Equivalent states will automatically be assigned the same combination of carry variables if γ-variable having the maximum number of subscripts are chosen wherever possible.

This statement can be proved by the following technique: it can be shown that γ-variables required to directly generate the outputs (call these γ^1) correspond to the sets of states which are indistinguishable by an experiment of length one.[5] The γ-variables required to generate the Γ^1 (which correspond to the γ^1) correspond to the sets of states indistinguishable by an experiment of length two, etc. In Table VIII, states 1 and 2, 3 and 4, 5 and 6, and 7 and 8 are equivalent since they have the same carry-variable assignments.

Sneak Paths

As mentioned previously, the assignment procedure just discussed does not guarantee that there will be no sneak paths in the resulting network. However, it is possible to determine their existence directly from the assignment table. Those discussed here involve only a single cell. Sneak paths involving several cells are possible, but occur rarely and must be checked by inspection of the network.

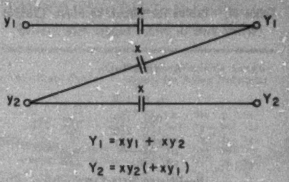


Fig. 6. Basic sneak path situation

The basic sneak-path situation is shown in Fig. 6. For this cell the desired functions are $Y_1=xy_1+xy_2$ and $Y_2=xy_2$. When the cell is designed, Y_2 is equal to xy_1+xy_2 instead of xy_2 because of a sneak path from y_1 to Y_1 to y_2 to Y_2. This sneak path cannot be eliminated without resorting to devices other than contacts, e.g., diodes. The specified functions require that when $x=1$, lead Y_1 is connected to lead y_2, and lead Y_2 is also connected to lead y_2. Thus, when $x=1$ leads, Y_1 and Y_2 must be connected to one another since they are both connected to the common lead y_2. From this it follows that when $x=1$, Y_1 and Y_2 must have the same signals and therefore that Y_1 $(x=1)$ $=Y_2$ $(x=1)$. This proves that it is impossible to design a contact network to have the specified Y_1 and Y_2 functions.

In general, the following rule applies:

> Whenever there is a term such as xy_i which appears in both of two output carry functions, any term such as xy_j which appears in only one of the functions will appear in both of the functions actually realized by a contact network.

For example, the functions for Table II—B are $Y_1=x'y_1+x'y_3$, $Y_2=xy_1$, $Y_3=x'$

Table IX. Tables Corresponding to the State Table of Table II

A—Inverse State Table

	x	
S_j	0	1
1	1	–
2	–	14
3	–	2
4	34	–
5	25	35
Z	34	34

s_j

B—Assignment Table

	x	
Γ	0	1
34	34	2
2	—	14
14	14, 34	—

Table X. Tables for Circuit to Have Output When Two or More Inputs Are Active

A—State Table

	x		
s_j	0	1	Z
0	0	1	0
1	1	2	0
2	2	2	1
	S_j		

B—Inverse State Table

	x	
S_j	0	1
0	0	–
1	1	0
2	2	1, 2
	s_j	

C—Assignment Table

	x	
Γ	0	1
Z	0	2
2	2	12
12	12	(012)

Table XI. State Table for Illustrating Periodic Carry-Variable Assignment

A—State Table

	x		
s_j	0	1	
1	4	3	0
2	3	4	1
3	1	3	0
4	2	1	1
	S_j		

B—Carry-Variable Assignment

s_j	v_1v_2		w_1w_2		y_1y_2	
1	1	1	0	0	0	0
2	1	0	0	1	1	1
3	0	0	1	0	1	0
4	0	1	1	1	0	1

Table XII. Carry Functions for Table XI

A—Functions with y-Variables Used as Carry Variables

$$Y_1=x'y_2+xy_2',\quad Y_2=x'y_1'+xy_1y_2,$$
$$Z=Y_{2n}=x_n'y_{1n}'+x_ny_{1n}y_{2n}$$

B—Functions with w-Variables Used as Carry Variables

$$W_1=w_1'+xw_2',\quad W_2=x'w_1w_2+xw_1'w_2+x'w_1'w_2'$$
$$Z=W_{2n}=x_n'w_{1n}w_{2n}+x_n'w_{1n}'w_{2n}'+x_nw_{1n}'w_{2n}$$

C—Function with v-Variables Used as Carry Variables

$$V_1=v_1'(x'+v_2),\quad V_2=(x+v_1+v_2')(x+v_1'+v_2)(x'+v_1+v_2)(x'+v_1'+v_2')$$
$$Z=V_{1n}'V_{2n}+V_{1n}V_{2n}'=(x_n+v_{2n}')(x_n'+v_{1n}'+v_{2n})$$

D—Functions with y- and w-Variables Used as Carry Variables

$$Y_1=x'w_2+xw_2',\quad Y_2=x'w_1w_2+x'w_1'w_2'+xw_1'w_2$$
$$Z=Y_{2n}=x_n'w_{1n}w_{2n}+x_n'w_{1n}'w_{2n}'+x_nw_{1n}'w_{2n}$$
$$W_1=(x+y_1'+y_2)(y_1+y_2'),\quad W_2=x'y_1'+xy_1y_2$$
$$Z=W_{2n}=x_n'y_{1n}'+x_ny_{1n}y_{2n}$$

E—Functions with y- and v-Variables Used as Carry Variables

$$Y_1=x'v_1v_2'+x'v_1'v_2+xv_1'v_2+xv_1v_2,\quad Y_2=x'v_2+xv_1v_2'$$
$$Z=Y_{2n}=(x_n'v_{2n}+x_nv_{1n}v_{2n}')$$
$$V_1=x'y_1y_2'+y_1'y_2,\quad V_2=x'y_2'+xy_2$$
$$Z=V_{1n}V_{2n}'+V_{1n}'V_{2n}=x_n'y_{1n}'+x_ny_{1n}y_{2n}$$

F—Functions with w- and v-Variables Used as Carry Variables

$$W_1=v_1+xv_2',\quad W_2=x'v_2+xv_1v_2',$$
$$Z=W_{2n}=x_n'v_{2n}+x_nv_{1n}v_{2n}'$$
$$V_1=w_1(x'+w_2),\quad V_2=(x+w_2')(x'+w_2),$$
$$Z=V_{1n}'V_{2n}+V_{1n}V_{2n}'=x_n'w_{1n}w_{2n}+x_n'w_{1n}'w_2'{}_n+xw_{1n}'w_{2n}$$

Table XIII. Derivation of Functions of Table XII

			x		
s_j	w_1	w_2	0	1	Z
1	0	0	01	10	0
2	0	1	10	01	1
3	1	0	00	10	0
4	1	1	11	00	1
			Y_1Y_2		

			x		
s_j	y_1	y_2	0	1	Z
1	0	0	11	10	0
2	1	1	10	11	1
3	1	0	00	10	0
4	0	1	01	00	1
			W_1W_2		

Table XIV. Equipment Required in Networks for Functions of Table XII with And-Gates and Or-Gates Used

Carry Variables	No. of Gates Per Cell	No. of Gate Inputs Per Cell	No. of Inverters Per Cell
Y_1Y_2, y_1y_2	6	13	2
W_1W_2, w_1w_2	6	15	2
V_1V_2, v_1v_2	7	19	2
Y_1Y_2, w_1w_2	7	17	2
W_1W_2, y_1y_2	6	14	2
Average	6.5	15.5	2
Y_1Y_2, v_1v_2	8	21	2
V_1V_2, y_1y_2	6	13	2
Average	7	17	2
W_1W_2, v_1v_2	5	11	1
V_1V_2, w_1w_2	4	8	1
Average	4.5	9.5	1

y_2+xy_2. Since $x'y_2$ appears in both Y_1 and Y_2, $x'y_1$, which appears in only Y_1, will appear in Y_2 and Y_1 in the actual network. This is precisely the sneak path pointed out previously. The term xy_2, which appears in Y_2 but not in Y_1, is not involved since it contains an x rather than an x'.

The rule just stated in terms of functions can also be applied directly to the assignment table:

> If in any column of an assignment table, two rows have any entries in common, then any entry which appears in either row must appear in both rows.

In the assignment table of Table IX—B, the entry 34 appears in both the first and third rows of the 0 column. Therefore the actual network will correspond to an assignment table in which the 14 entry also appeared in both the first and third rows. This agrees with the preceding discussion in which $\gamma_{14}=y_1$, $\gamma_2=y_2$, and $\gamma_{34}=y_3$, and the sneak path was $x'y_1=x'\gamma_{14}$ in $Y_3=\Gamma_{34}$. For this particular table and assignment it is not possible to avoid the sneak path. However, for some tables and assignments it is possible to remove sneak paths by adding one or more extra contacts in the cell.

Table X shows a state table and an assignment, $Y_2=\gamma_2$, $y_1=\gamma_{12}$, for which the resulting cell contains a sneak path. The functions $Y_1=y_1+x$, and $Y_2=y_2+xy_1$ can be written as $Y_1=x(1+y_1)+x'y_1$ and $Y_2=x(y_2+y_1)+x'y_2$. For this table $\gamma_{012}=1$, since the network is always in one of its three states. Since they both contain the term xy_1, the resulting network will produce a sneak path of x in Y_2 and of xy_2 in Y_1. The x term in Y_2 cannot be

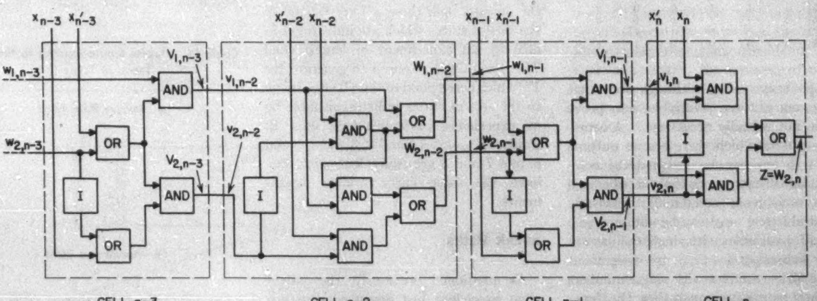


Fig. 7. Network corresponding to Table XII—F

Table XV. State Table and Corresponding Functions

A—State Table

	x 0	1	Z
sj			
1	4	3	0
2	3	4	1
3	1	2	0
4	2	1	1
	Sj		

B—Assignment Table

y_1	y_2	x 0	1	Z
0	0	01	10	0
1	1	10	01	1
1	0	00	11	0
0	1	11	00	1
		Y_1Y_2		

C—Functions

$$Y_1=x'y_2+xy_2'$$
$$Y_2=x'y_1'+xy_1$$
$$Z=Y_2$$

allowed. By setting $Y_1=x+x'y_1=x(1)+x'(y_1)$ and $Y_2=y_2+xy_1=x(y_1+y_2)+-x'(y_2)$, the common term xy_1 is removed and the sneak path destroyed.

The possibility arises that there may be some state tables which cannot be realized by a contact network because none of the possible carry-variable assignments correspond to networks which can be realized without unallowed sneak paths. This possibility can be immediately discarded since, for any state table a one-hot assignment will correspond to a realizable contact network. (For some state tables, this is the only realizable assignment; see the discussion of Table IV.) The fact that a one-hot assignment produces functions which contain no products or complements of carry-variables has been discussed. It is thus only necessary to demonstrate that a one-hot assignment never introduces sneak paths.

For a one-hot assignment, the assignment table is identical with the inverse state table. In the inverse state table each number appears only once in each column since any input-carry state can lead to only one output-carry state for each input combination. Since each number appears only once in a column, it is impossible for two rows to contain the same number in any column. Thus there can be no sneak paths since the criterion for a sneak-path is that two rows of the assignment table have a common entry in some one column.

Varying Assignments

In the preceding discussion of carry-variable assignments it has been assumed that the assignment of carry variables to states was constant throughout the network. This is necessary in sequential circuits, but for iterative circuits it need not be so, and it may be possible to obtain a more desirable network by varying the correspondence between states and carry variables. In particular, it may be possible to reduce the complexity of the cells and sometimes also to reduce the required number of carry leads. Of course, the network will no longer consist of a chain of identical cells; only every other cell or every third cell, etc., may be identical. This sacrifice of complete uniformity may be worth while if a sufficient saving is realized in the complexity of the individual cells.

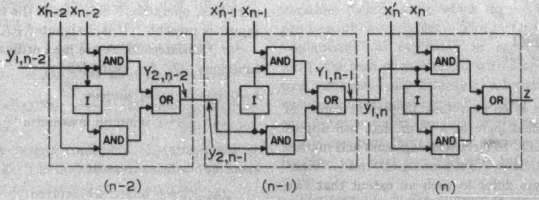
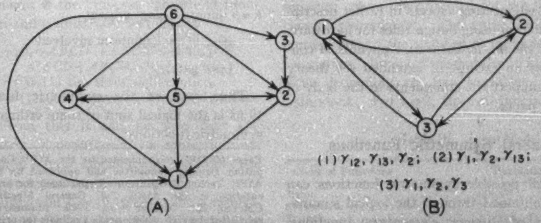


Fig. 8. Network for Table XV



(1) γ_{12}, γ_{13}, γ_2; (2) γ_1, γ_2, γ_{13};

(3) γ_1, γ_2, γ_3

Fig. 9. Diagram showing which cells with different input-carry and output-carry variable assignments are realizable with a relay network. A—For Table XVI. B—For Table XVII

Table XVI. State Table for Network Having an Output with 0, 1, or 3 Inputs Active

A—State Table

	x 0	1	Z
sj			
1	1	2	1
2	2	3	1
3	3	4	0
4	4	5	1
5	5	5	0

B—Carry-Variable Assignments for Relay Networks

1.	γ_{523}	γ_{52}	γ_1	γ_5
2.	γ_{513}	γ_3	γ_1	γ_5
3.	γ_{53}	γ_2	γ_1	γ_5
4.	γ_{13}	γ_{53}	γ_1	γ_5
5.	γ_{13}	γ_5	γ_2	γ_1
6.	γ_3	γ_2	γ_1	γ_5

Table XVII. Illustrative State Table

	x 0	1	Z
sj			
1	1	2	1
2	1	3	1
3	4	2	0
4	4	4	0

The most economical iterative network (in number of gates per cell) for the state table shown in Table XI—A is one in which two carry-variable assignments alternate throughout the network. Since this table contains four carry states, there are three "different" assignments to two carry variables.[4] These assignments are shown in Table XI—B, where the variables v_1v_2 are used for one assignment, w_1w_2 for another, and y_1y_2 for the third. The functions which result from these assignments are shown in Table XII—A, B, and C.

The functions of Table XII—D, E, and F result when one pair of variables (e.g., y_1y_2) are used as input-carry variables and another pair of variables (e.g., V_1V_2), are used as output-carry variables. Table XIII illustrates the derivation of the functions of Table XII—D, and Table XIV lists the equipment required by networks for the functions of Table XII. Clearly, the most economical network is obtained by using w_1w_2 and v_1v_2 as carry variables. The network which results from this choice of carry variables is represented in Fig. 7.

Table XV shows a state table for which it is possible to reduce the number of carry leads from two to one by using different carry leads for different cells. The corresponding network is shown in Fig. 8. In this network there is only one carry lead between cells, but this carry lead sometimes represents y_1 and sometimes y_2. This is possible because for this particular state table and assignment $Y_2(=Z)$ is not a function of y_2, and Y_1 is not a function of y_1. Thus, a cell requires only y_1 or y_2 as an input carry. This is not true for other carry-variable assignments. For example, if the assignment $y_1=\gamma_{34}$, $y_2=\gamma_{24}$ is chosen the functions are:

$Z=Y_2,\ Y_1=y_1'$

$Y_2=x'(y_1y_2+y_1'y_2')+x(y_1y_2'+y_1'y_2)$

These functions require two carry leads per cell and would have more complicated cell circuitry than that shown in Fig. 8.

It is also possible to design relay networks for which the carry-variable assignment varies within the network. However, the same problems of realizability which were discussed in connection with networks having constant carry-variable assignments arise. It is not always possible to construct a network in which two assignments alternate even if each of the assignments corresponds to a realizable network with a constant carry-variable assignment. For example, there are six carry-variable assignments for Table XVI which correspond to physically realizable carry functions. It is not possible to construct a cell of relay contacts which has the assignment 1 for its input carries and assignment 2 for its output carries.

Fig. 9(A) is a diagram which shows which cells with different input-carry variable and output-carry variable assignments are realizable. An arrow from mode i to node j indicates that it is possible to construct a cell with assignment i for the input carries and assignment j for the output carries, and the absence of an arrow indicates that such a cell is impossible. This diagram shows that it is not possible to construct a relay network in which any pair of assignments alternate. In this case, it is possible only to construct a network in which the carry-variable assignment changes from one assignment to another (e.g., from point 6 to point 1) and then remains constant throughout the remainder of the network.

State tables do exist for which relay networks are possible in which two (or more) carry-variable assignments alternate. An example of such a state table is shown in Table XVII. Fig. 9(B) is a diagram which shows for Table XVII which cells are possible with different input-carry and output-carry assignments.

There is no method known for determining directly from the state table whether it is advantageous to vary the correspondence between carry variables and states within a network. It is necessary to derive the appropriate carry functions (as in Table XII) in order to assess the advantages of such a scheme for a given state table.

Summary

The problem of assigning combinations of carry variables to the network states has been discussed and procedures for obtaining assignments corresponding to physically realizable contact networks have been presented, with the possibility of varying assignments within the network also considered.

References

1. THE SYNTHESIS OF SEQUENTIAL SWITCHING CIRCUITS, D. A. Huffman. *Journal*, Franklin Institute, Philadelphia, Pa., vol. 257, nos. 3 and 4, Mar. and Apr. 1954.

2. SWITCHING CIRCUITS AND LOGICAL DESIGN (book), S. H. Caldwell. John Wiley & Sons, Inc., New York, N. Y., 1958.

3. A COMPARISON OF SEQUENTIAL AND ITERATIVE CIRCUITS, E. J. McCluskey, Jr. *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 78, 1959 (Jan. 1960 section), pp. 1039–44.

4. A NOTE ON THE NUMBER OF INTERNAL VARIABLE ASSIGNMENTS FOR SEQUENTIAL SWITCHING CIRCUITS, E. J. McCluskey, Jr., S. H. Unger. *Transactions*, Professional Group on Electronic Computers, Institute of Radio Engineers, New York, N. Y., vol. PGEC-8, Dec. 1959, pp. 439–40.

5. GEDANKEN–EXPERIMENTS ON SEQUENTIAL MACHINES, E. F. Moore. "Automata Studies," no. 34, "Annals of Mathematics Studies," C. E. Shannon, J. McCarthy, editors. Princeton University Press, Princeton, N. J., 1956, pp. 129–53.

Universal Modulus

Matrix Logic VI

E. J. SCHUBERT
ASSOCIATE MEMBER AIEE

Synopsis: This paper continues the series based on the use of matrices in the design of switching networks. To avoid redundancy within the series, all previous papers on matrix logic are considered as prerequisite, in particular, references 1 and 2. This paper establishes a general classification scheme for logical functions, as follows: 1. General symmetric functions, which are the sum of ordinary symmetric functions. 2. Trivial and redundant functions. 3. Universal functions, which are capable of performing any logical functions by appropriate iterations.

THE LOGICAL or Boolean functions of binary variables may be classified in various categories. The purpose of the classification is to determine those logical functions which are independently capable of serving as the only building blocks in any logical network. The mechanization of these logical functions may be designated as "universal modulus."

It is no easy task to present topics of applied logic in a form that will appeal equally to electrical engineers and mathematicians. Terminologies and viewpoints differ to such an extent that two completely independent approaches must be taken. Therefore, for the electrical engineer, the paper proper emphasizes the engineering aspects in rather descriptive form using design rules for hardware. For the mathematician, Appendix II contains an excerpt of extended set theory to support the statements in the body of the paper.

General Symmetric Functions

All possible symmetric functions can be obtained through the logical summation of distinct elementary functions, which are called "ordinary symmetric." By definition, "a symmetric function ${}_pO_k$ is ordinary if it is operative for any combination of p inputs taken k at a time."[2] A typical ordinary symmetric function is the "half-adder" or "exclusive OR" circuit, which is a circuit of two input variables, operative if only one of the two inputs is present. It is designated ${}_2O_1$.

For p variables, there are $p+1$ ordinary functions ${}_pO_k$, for example:

$k=0$ inputs being present

${}_pO_0$: $\left\|\begin{matrix}p\\0\end{matrix}\right\|=1$ minterm involved

(NOR gate).

$k=1$ inputs being present

${}_pO_1$; $\left\|\begin{matrix}p\\1\end{matrix}\right\|=p$ minterms involved

.

.

.

$k=p$ inputs being present

${}_pO_p$; $\left\|\begin{matrix}p\\p\end{matrix}\right\|=1$ minterm involved

(AND gate).

Thus, any of the symmetric functions is the logical sum of some ordinary symmetric functions:

Paper 60-1181, recommended by the AIEE Computing Devices Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Chicago, Ill., October 9–14, 1960. Manuscript submitted July 15, 1959; made available for printing July 28, 1960.

E. J. SCHUBERT is with Monitor Systems, Inc., Fort Washington, Pa.

Table I. Classification of 2-Variable Functions

$$Y=H_0X_1'X_2'+H_1X_1X_2'+H_2X_1'X_2+H_3X_1X_2=\underset{i}{U}H_i\cdot N_i=(h)\cdot(n)$$

Conventional	Operator (h)	Symmetric (s)	Redundant r=	Universal	Circuit
$Y_0=0$	(0000)	$(\underline{0}00)$	1, 2		OFF
$Y_1=X_1'X_2'$	(1000)	$(\underline{1}00)$		X	NOR
$Y_2=X_1X_2'$	(0100)			X	
$Y_3=X_2$	(1100)		1		
$Y_4=X_1'X_2$	(0010)			X	
$Y_5=X_1'$	(1010)		1		
$Y_6=X_2'X_1+X_2X_1'$	(0110)	$(\underline{0}10)$			HALF ADDER
$Y_7=X_1'+X_2$	(1110)	$(\underline{1}10)$		X	SHEFFER
$Y_{10}=X_1X_2$	(0001)	$(\underline{0}01)$			AND
$Y_{11}=X_1X_2+X_1'X_2'$	(1001)	$(\underline{1}01)$			EQUALITY
$Y_{12}=X_1$	(0101)		1		
$Y_{13}=X_1+X_2'$	(1101)			X	
$Y_{14}=X_2$	(0011)		1		
$Y_{15}=X_2+X_1'$	(1011)			X	
$Y_{16}=X_1+X_2$	(0111)	$(\underline{0}11)$			OR
$Y_{17}=1$	(1111)	$(\underline{1}11)$	1, 2		ON

Table II. Number of Redundant Functions ${}_pQ_r$ and Probability of Redundancy P(p)

	${}_pQ_r$					
$r=$	1	2	3	4	5	$P(p)$
$p=1$	2	.	.	.		0.5
2	6	2	.	.		0.375
3	38	8	2	.		0.149
4	918	46	10	2		0.014
5	325, 382	2,292	112	12	2	$<10^{-6}$

$$Y=\bigcup_{k=0}^{p}{}_pS_k\cdot{}_pO_k \qquad (1)$$

In equation 1, the composite symmetric function Y is said to be the logical sum U (union) of the dot products (logical product) of all $p+1$ possible ordinary switching functions ${}_pO_k$ with an associated coefficient ${}_pS_k$. The latter is binary and may be expressed by a row matrix (s). As an example, the ordinary switching function of three out of four inputs is shown.

$$ABCD'+ABC'D+AB'CD+A'BCD$$
$$={}_4O_3\left\|\begin{matrix}4\\3\end{matrix}\right\|=4 \text{ minterms involved.}$$

$(s)=(\underline{0}0010)$; ${}_4S_3=1$; all other coefficients are zero.

The matrix coefficient ${}_4S_0$ corresponding to $k=0$ is underlined to define the location of column 0 in the matrix.

A function being true for at least k inputs is a threshold function. Its complement is the "inverse" threshold function and is illustrated as follows:

$${}_4O_3+{}_4O_4=ABCD'+ABC'D+AB'CD+A'BCD+ABCD.$$

$(s)=(\underline{0}0011)$ the threshold function.
$(s)'=(\underline{1}1100)$ the inverse threshold function.

From this it can be seen that the matrix (s) describes any symmetric function in a more efficient way than does the conventional function matrix.[2]

Equation 1 lends itself to the computation of the total number of symmetric switching functions for any number of p variables. Matrix (s) is a binary word comprising $p+1$ bits. Accordingly there can be only 2^{p+1} different numbers or matrices of (s) or different symmetric functions. The relative number of symmetric functions decreases rapidly when there are an increasing number of the p input variables.

Redundant and Trivial Functions

All possible switching functions of p variables include some which do not depend on all p variables. The latter are "redundant" if expressed in canonical form (the sum of minterms) with respect to p variables. By definition, a function eventually becomes "trivial" if only one significant variable remains, or the function is constant, 0 or 1. For a nontrivial function, at least two variables must remain.

The term ${}_pO_r$ is the number of switching functions of p variables wherein r variables are redundant. In Table I, a separate column is provided for the feature of redundancy. The value 1 or 2 corresponds to the number of redundant r variables. If $r=1$, then 6 of the 16 functions depending on two variables are redundant. For $r=2$, the two functions Y_0 and Y_{17} are redundant and constant. Actual computation of the number of redundant switching functions is treated in Appendix I. In Table II, the results indicate that the probability $P(p)$ of a function to be redundant decreases rapidly with increasing number of p variables. In the case of four variables, approximately 1 out of 50 functions exhibits redundancy. Around this minority centers the major effort of reduction or minimization techniques.

Universal Functions

Any logical function is a logical operator, because it performs an operation on the inputs. Diode logic comprising the AND gate and the OR buffer, requires an inverter to build up any logical function. In this concept, there are three operators involved: AND, OR, and NOT.

For some time, certain "primitive" functions have been known, capable of performing any logic.[3] It is suggested, however, that because of their outstanding capabilities the term "universal" be adapted for such functions. To qualify as universal, a function must have at least two inputs and be capable of inverting for at least one input. These are necessary but not sufficient conditions as far as symmetric functions are concerned. The additional criterion for symmetric functions can be stated with reference to the matrix (s), in which the entries 2 indicate irrelevant, i.e., "don't care" coefficients which might be either 0 or 1.

$$(s)=(\underline{1}222\ldots2220) \qquad (2)$$

For $p=2$, the following symmetric functions will satisfy these conditions:

$(s)=(\underline{1}00)$ NOR gate
$(s)=(\underline{1}10)$ SHEFFER'S STROKE

It will be noted that there are no other symmetric functions of two variables which are simultaneously universal.

From equation 2, the number of symmetric universal functions becomes immediate because, in this matrix, the p-1 binary coefficients can be permuted in 2^{p-1} ways. The first and last coefficients are constant, 1, and 0 respectively.

The number of asymmetric functions a_6 which are universal can be predicted from the number of trivial and symmetric

Table III. Input Variables Versus Population of Classes

Class of Functions	Population	$p=2$	$p=3$	$p=4$
Total	$a_1=2^{2^p}$	16	256	65, 536
Symmetric	$a_2=2^{p+1}$	8	16	32
Trivial	$a_3=2(p+1)$	6	8	10
Redundant	$a_4={}_pQ_1$	6	38	918
Symmetric-universal	$a_5=2^{p-1}$	2	4	8
Asymmetric-universal	$a_6=a_1-a_2-a_3+2$	4	234	65, 496

functions. Two of the trivial functions are symmetric at the same time. Referring to the entries in Table III,

$$a_4 = a_1 - a_2 - a_3 + 2 \tag{3A}$$
$$= 2^{2^p} - 2^{p+1} - 2(p-1) \tag{3B}$$

In Table I the complete set of functions of two variables is listed using the function matrix (h) as reference.[2]

UNIVERSAL MODULUS

It appears highly desirable to restrict the different types of building blocks for efficient production, assembly, and maintenance of switching systems. The eventual choice will mostly be based upon the over-all reliability of the circuit performing the switching function, and upon economical considerations.

The mechanization of the universal function is called "universal modulus." According to basic theorems any logical network can be assembled with the use of a single type of universal modulus. Even if the universal modulus comprises only two inputs, a cascade of these moduli will accommodate any larger number of inputs. The technique for the logical design has been developed previously and is suitable for any type of logical operator.[1]

Symmetric universal moduli are preferred in the present trend to solid-state circuits. Transistor and magnetic logic utilize the NOR function extensively. Inverse threshold functions are possible for reliable operation, but somewhat controversial because of their extremely close tolerances for components.

One disadvantage of universal moduli is that conventional Boolean algebra cannot be modified to handle such operators efficiently. Therefore, the engineering effort becomes more pronounced and costly, particularly for custom-built systems where neither computer routines nor improvements by redesign are applicable.

UNIVERSAL PACKAGE

In some cases, a circuit can be developed for a modulus which provides more than one logical output. If at least one combination of the logical functions yields a universal function, a circuit can be considered a universal package, even if each of the logical functions is not universal by itself.

The advantage of the universal package over the simple universal modulus with only one function is that greater simplicity of the final network can be gained with less hardware. Typical for such a package is the combination of an AND gate and an inverter. Referring to Table I, each of the outputs correspond to Y_3 and Y_{10} respectively. Taking two of the three inputs, with appropriate connection of the remaining input to one output, yields three different universal functions Y_2, Y_4, and Y_7.

UNIVERSAL MODULI FOR SEQUENTIAL LOGIC

Combinational logic can be performed in passive networks, provided the signal sources feeding into the inputs have sufficient power. In sequential logic, active elements are essential to maintain or retrieve the information from the storage. A basic storage element and its associated logic are described by a transformation derived previously.[4] To set a storage cell, a trigger signal $\dot{Z}$ must be replaced by a continuous feedback signal Z.

Two cascaded universal moduli, with a feedback loop from the output of the second modulus into an input of the first, will exhibit the latching function of a flip-flop. In general, any closed loop containing an even number of active symmetric universal moduli will perform a latching or memory function.

$$Y = (Y')' \ldots \text{two inverters} \tag{4A}$$

$$Y = (((Y')')')' \ldots \text{four cascaded inverters} \tag{4B}$$

In a subsequent paper "Logical Design by Regression—Matrix Logic VII," a novel design routine will be discussed which is highly efficient and directly applicable for any arbitrary universal function.

Conclusions

A classification for logical functions and the determination of their numbers indicates the pronounced majority of asymmetric and universal functions.

Existing hardware lends itself to a preference for symmetric universal moduli. The difficulties encountered in the logical design using these moduli, are due to basic limitations of Boolean algebra.

Appendix I. Number of Redundant Functions

Definition: ${}_pQ_r$ is the number of redundant functions in which at least r of the p variables are irrelevant $0<r<p$.

1. ${}_pQ_p = 2$; if all p variables are irrelevant, then 0 and 1 are the only remaining functions.

2. If k variables are irrelevant, then the p variables may be taken p-k at a time. The number of combinations are:

$$\left\|\begin{matrix} p \\ p-k \end{matrix}\right\| = \left\|\begin{matrix} p \\ k \end{matrix}\right\|$$

3. Any set of p-k variables can form $2^{2^{p-k}} - {}_{p-k}Q_1$ distinct nonredundant functions, excluding those wherein at least one of the p-k variables is redundant.

All sets of redundant functions of p-k variables will intersect in a proper subset containing ${}_pQ_{k+1}$ redundant functions characterized by $k+1$ irrelevant variables.

The following relation is immediate:

$${}_pQ_k = {}_pQ_{k+1} + \left\|\begin{matrix} p \\ k \end{matrix}\right\| \left(2^{2^{p-k}} - {}_{p-k}Q_1\right)$$

Substituting $r=k+1$ yields the recursion formula:

$${}_pQ_{r-1} = {}_pQ_r + \left\|\begin{matrix} p \\ r-1 \end{matrix}\right\| \left(2^{2^{p-r+1}} - {}_{p-r+1}Q_1\right) \tag{5}$$

The initial steps for computing Table II are:

$${}_1Q_1 = 2$$

$${}_2Q_2 = 2$$

$${}_2Q_1 = 2 + \left\|\begin{matrix} 2 \\ 1 \end{matrix}\right\| (2^2 - 2) = 6$$

$${}_3Q_3 = 2$$

$${}_3Q_2 = {}_3Q_3 + \left\|\begin{matrix} 3 \\ 2 \end{matrix}\right\| (2^2 - {}_1Q_1)$$
$$= 2 + 3(4-2) = 8$$

$$2Q_1 = {}_3Q_2 + \left\|\begin{matrix} 3 \\ 1 \end{matrix}\right\| (2^{2^2} - {}_2Q_1)$$
$$= 8 + 3(16-6) = 38$$

Appendix II. Criteria of Universal Laws of Composition

The terminology in this derivation follows closely the text of Jacobson's lectures.[5]

Definition: A Boolean universal algebra is a lattice with 0 and 1 restricted to a single law of composition, which is complementing.

Theorem: A universal law of composition in a Boolean universal algebra is complementing and excludes functions related to symmetric difference.

Corollary: A universal law of composition is complementing and either nonassociative (asymmetric), or if associative (symmetric) then not related to symmetric difference. (Equivalence is the complement of symmetric difference and therefore excluded).

Proof:

1. A Boolean algebra is a lattice with 0 and 1 that is distributed and complemented.
2. All Boolean algebras are isomorphic.
3. Boolean algebra and Boolean ring with identity are equivalent.
4. The composition's symmetric difference, intersection ($\oplus$,.), and the lattice are a Boolean ring. Therefore, neither symmetric difference nor intersection by itself suffices as single law of composition in a Boolean universal algebra.

Symmetric difference is identical with binary addition modulo 2 as mechanized in half-adders, $A'B + AB'$. Its complement is the equivalence, $A'B' + AB$. Both are the symmetric functions as excluded by the theorem.

References

1. Matrix Synthesis of High-Speed Logic, E. J. Schubert. *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 78, Mar. 1959, pp. 4–8.

2. Symmetric Switching Functions—Matrix Logic V, E. J. Schubert. *Ibid.*, 1959 (Jan. 1960 section), pp. 1083–97.

3. Theory of Logical Nets, A. W. Burks, T. B. Wright. *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 41, no. 10, Oct. 1953, p. 1357.

4. Matrix Algebra of Sequential Logic—Matrix Logic III, E. J. Schubert. *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 78, 1959 (Jan. 1960 section), pp. 1074–79.

5. Lectures in Abstract Algebra (book), N. Jacobson. D. Van Nostrand Company, Inc., Princeton, N. J., 1951.

Teaching the Uses of Digital Computers at the Technical University of Denmark

OLE IMMANUEL FRANKSEN
STUDENT MEMBER AIEE

Synopsis: A method for teaching electrical engineering students the use of digital computers in solving electric machinery problems, especially with regard to iterative methods, is explained. By emphasizing the process of stating actual problems and formulating them in flow diagrams, more general principles are brought to light. The method is illustrated with an example of a power transformer winding design. The use of the international computer language Algol, is demonstrated and discussed.

A RESEARCH PROGRAM investigating the use of digital computers in the field of solving engineering and design problems was initiated in the spring of 1957 in the Electric Machinery Department (headed by the Dean of Electrical Engineering, Prof. L. Hyldgaard-Jensen) at the Technical University of Denmark. The main purpose was to study this area from an educational point of view and to try, if possible, to develop a general approach which could be taught to electrical engineering students.

As we were especially interested in the problems of the iterative type (e.g., design), it was decided to develop a program for designing power transformers on the Danish computer Dask, to gain the necessary experience. This program (approximately 12,000 computer instructions) was finished during the summer of 1959. Before the completion of the transformer program, the new international computer language, Algol, had advanced so far that its use was incorporated into the investigation.

Using a notation close to normal mathematical standards, Algol describes the computational processes to be done in certain arithmetical and nonarithmetical statements (in English), each statement given a single, precise meaning. When the computer receives the Algol instructions, it translates these into its own terms by means of a special program called a "compiler."[1]

We have found that it is not yet possible to teach general procedures by lectures, enabling the student to formulate logically the problems in a flow diagram, even though it is possible to state part of the techniques in general terms. For this reason (which, by the way, means that the student only "learns by doing") we have not prepared a special computer course, but have incorporated an introduction to the use of digital computers in the graduate course in electric machinery.

The purpose here is to describe how we present this subject to our students, and what experience we have gained in doing so.

Scope

To secure a master's degree in electrical engineering at the Technical University of Denmark (which does not offer a bachelor's degree) requires normally $5^1/_2$ years' study, the last half-year being spent on a thesis. The "Computer Introduction" is given at the beginning of the student's last year, after he has specialized in electric machinery and has become somewhat familiar with the problems.

From the day the student enters the University, he is encouraged to take an optional course in programing (until 1959 in the Dask machine code and since then in Algol). It should be emphasized that the Introduction is intended to give the student the ability to so formulate the problem that it can be programed for the computer by personnel who know nothing at all about the problem itself. We feel that it is very important to familiarize the student with programing methods, enabling him to see when it will be feasible for him to put his problem on a digital computer. In practice, the computer specialist usually is unable to formulate a specialized problem for the computer, so the engineer must know how to do this.

The Introduction is therefore directed not toward future computer specialists but to engineers who are interested in the general utility of digital computers from the point of view that they want to be able to use the computers not only as large-scale slide rules. This requires, of course, quite a lot of experience, part of which the student may obtain during the work on his thesis.

Some of the most difficult problems to formulate for the computer are those of design, which the engineer solves by an intuitive cut-and-try method. By taking examples with backgrounds familiar to the student, we show him how to formulate them for the computer, i.e., what data were originally available; how these data were prepared; and how one must forget the usual methods and look at the problem anew (cf., the problems of the pressboard spacers and the computation of the winding temperature in the example which will be treated later). After having presented each problem along these lines, we encourage the student to criticize the solution and to suggest improvements.

Another important purpose of the Introduction is to emphasize the need of documenting the solution of the problem. Lack of such documentation often leads to computer programs that are too expensive to keep up to date, or even useless, owing, for example, to changes in personnel.

A very good and clear way of documenting this information is to work out flow diagrams. By such means, especially with involved problems, one is also able to eliminate the logical errors which usually occur when the solution to a problem is attempted mentally.

Flow Diagraming

Theoretical Considerations

The two properties of the digital computer which principally determine the method of attack in putting a problem on it are:

Paper 60-1184, recommended by the AIEE Computing Devices Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Chicago, Ill., October 9–14, 1960. Manuscript submitted June 6, 1960; made available for printing September 2, 1960.

Ole Immanuel Franksen is with Rensselaer Polytechnic Institute, Troy, N. Y., on leave of absence from the Technical University of Denmark.

1. Arithmetically, the computer can do only the four basic operations: addition, subtraction, multiplication, and division.
2. Logically, the computer can compare only two numbers at a time and continue the series of computations along one of two alternative ways, depending on the result of the comparison.

A flow diagram can be defined as a map over the routes the computer has to follow to carry through the desired computations. This definition implies that the flow diagram, independently of the special computer used, should give a pictorial representation of the problem, stating the exact sequence of operations on the data and how the alternatives fit together.

Depending on the type of problem to be solved, there are, of course, many ways to work out a flow diagram, although they all basically consist of a collection of lines, arrows, and boxes, the last-mentioned containing instructions about the things to be done.

Normally, it is beneficial to work out a series of flow diagrams ranging from the bird's-eye view ("rough" flow diagram) to the ready-to-code diagram, not only from the point of view of getting a better solution but also because such a series is one of the best ways of transferring the solution to other personnel.

If the problem to be solved involves a long series of straightforward calculations (which, for example, is the case in the computation of performance), it may be advisable to list these equations in an appendix to the flow diagram and mark on this "Eqs. *n* to *p*," etc. If, on the other hand, the algebraic sequences between the different alternatives are relatively short (as in the design example to be treated later), then it may be better to have all the information on the flow diagram.

No matter how it is constructed, it is of vital importance that the flow diagram be as easy as possible to understand. It has been found desirable in working out a flow diagram to state the following:

1. Terminology.
2. Loops.
3. Numerical conditions.
4. Order.
5. Examples.

An explanation of items 1 and 2 follows immediately, while items 3, 4, and 5 will be discussed in the section "Practical Considerations."

A great help in clarifying the flow diagram is a standardized system of symbols and ways of writing the instructions. For the latter we adopted the ALGOL notation (as illustrated in Fig. 4) and expanded it with a short explanation, sometimes also adding a unit. These instructions plus comments are put into boxes which are numbered for identification. To indicate in which order the boxes must be run through during the computation, they are connected with lines, each of which is supplied with an arrow where it runs into another box or into a junction for several lines.

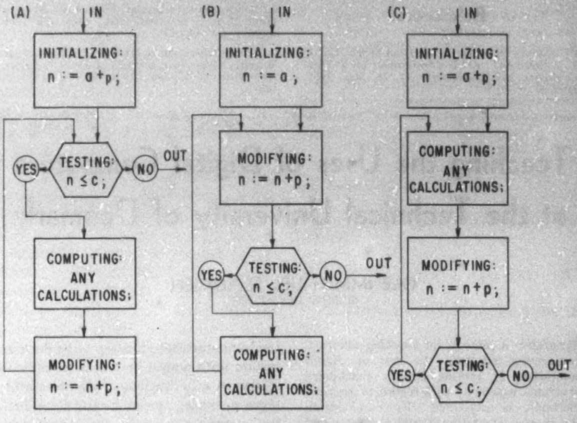


Fig. 1. Loops in computing

Because there is not as yet any international standard for the symbols, we do not insist on using a special set of symbols but explain to the student that it is normally sufficient to differentiate between four basic types of boxes, namely:

1. Input and output boxes.
2. Operation boxes.
3. Choice boxes.
4. Error boxes.

A particular flow diagram can contain more than one input and one output box, each of which has a function to define the entrance or exit with the name or a number, and at the same time it should list all the actual input or output parameters. It is advisable also to list the units of the parameters in order to prevent interconnection errors between several flow diagrams. The operation boxes describe the arithmetic computations which have to be done. The choice boxes are the only type with two exits, i.e., depending on a question which can be answered with YES or NO, they direct the computation along one of two alternative ways. The error boxes are used to indicate numbered stops which may be incorporated into the program.

Nearly always, one of the major factors in preparing a problem for solution on a digital computer is its limited memory space. Hence, it is necessary on the flow-diagraming level to consider where it is possible to use computation loops; by which is meant that parts of the computer program can be repeated with modifications. The savings in memory space and programing time thus obtained will always justify the longer machine time, and we shall therefore briefly treat the theory of loops in computing.

Referring to Fig. 1, it is always possible to divide a loop into four parts:

1. Initializing.
2. Testing.
3. Computing.
4. Modifying.

Interchange of the last three is freely possible. In practice, this interchange leads to only three different principles, as Fig. 1 implies, because the mutual order of parts 3 and 4 is unimportant for the function of the loop when they follow each other. It is hard to state a general rule to follow for the application of these three principles, which are so similar in theory but so different in practice. Which one to choose in a particular case depends not only on the computations which should be executed in the loop but also, for example, on outer loops.

Two questions must always be considered immediately after the construction of a loop:

1. Do the modification and the counter-constant work in the right way?
2. Is anything computed under the head-

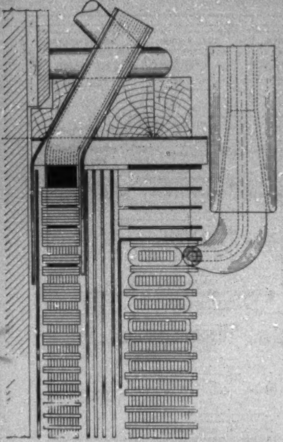


Fig. 2. Cross section of a pair of transformer windings

ing "Computation" which could just as well be computed outside the loop?

Considering question 1, it is always a good idea to see what would happen if the loop were to be executed only once. In an affirmative case, question 2 obviously involves a waste of machine time.

PRACTICAL CONSIDERATIONS

Numerical analysis is taught during the student's second year at the University, as a part of the applied mathematics course. Therefore, in the Introduction, we briefly consider only three different numerical relations, which, experience has shown, always amaze the student. These three sets of relations concern:

1. The floating decimal point.
2. The limited length of the accumulator.
3. The numerical accuracy of the original data.

All calculations coded in a programing language such as ALGOL are automatically performed with a floating decimal point. A repetitive type of problem in design is the determination of engineering standard values when these vary in discrete, often nonequidistant steps. We therefore try to tell the student (in ALGOL notation) what to do when the calculation gives as a result a nonstandard value, which has to be rounded off in a special way to a standard value. We also discuss the problem of rounding off and calculating with integers which have to be odd, even, or divisible by some other integer (i.e., the example in Fig. 4 where the ALGOL function "entier" is used for this purpose in the complex around box no. 38).

To illustrate the type of problems introduced by the limited length of the accumulator in the computer, some examples are given. It is shown:

1. How to arrange the order of a computation to avoid taking the difference between numbers of nearly equal size.
2. That there might be a difference between dividing with a number or multiplying by its reciprocal.
3. That sometimes a better result might be obtained by using Horner's diagram (also called the method of "nesting"), when values of a power series are being calculated, instead of using the built-in ALGOL "exponentation brackets."
4. How important it might be to have the different test values provided with tolerances; the problem of arranging the order of the tests, which is incidentally one of the major questions in design problems, is unfortunately a matter of experience.

Owing to the great numerical capabilities of the digital computer, one is often inclined to demand a number of digits after the decimal point, which is completely unrealistic from a physical point of view. The student is therefore urged to analyze carefully the original data and the problem from this aspect.

As just one error might be sufficient to ruin a whole program, it is obvious that good order and accuracy are of vital importance. Experience has shown that the number of errors grows with some power of the length of the program; this is why a good tactician always tries to subdivide his program into smaller procedures of limited length. As a rule of thumb, we advise the student to work out each of his flow diagrams on a single sheet of paper; this determines the maximum length of the individual procedures. Another aspect of this subdivision is that it helps the student to recognize repetitive problems and to work them out as routine (library) procedures. Good order also implies planning of input forms, output forms, and synonymous names for the parameters and an up-to-date and accurate correction of errors in the flow diagrams.

It is stressed that no problem can be regarded as finished before numerical test examples are worked out in such a way that every route in the flow diagrams can be run through and tested during the debugging process. However, owing to the iterations involved, the nature of a design problem is often such that it is impossible to work out a complete numerical example. In these cases, the student is advised to use the following techniques, which are very useful when the program is coded in ALGOL:

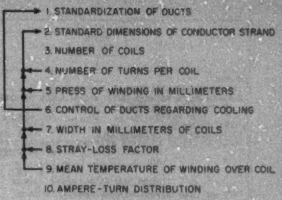

1. STANDARDIZATION OF DUCTS
2. STANDARD DIMENSIONS OF CONDUCTOR STRAND
3. NUMBER OF COILS
4. NUMBER OF TURNS PER COIL
5. PRESS OF WINDING IN MILLIMETERS
6. CONTROL OF DUCTS REGARDING COOLING
7. WIDTH IN MILLIMETERS OF COILS
8. STRAY-LOSS FACTOR
9. MEAN TEMPERATURE OF WINDING OVER COIL
10. AMPERE-TURN DISTRIBUTION

Fig. 3. Rough flow diagram of high-voltage disk coil winding

1. Divide the program into procedures, each of which can be easily supplied with numerical test examples for debugging.
2. Work out test examples for each of these procedures.
3. Prepare input parameters for the program as a whole.

The reason for this procedure is the sequential feature of ALGOL, which makes it rather easy to construct a tracing program which (during the tracing) will print out the output parameters of a procedure every time its exit is passed. Such a feature was planned to be incorporated in the ALGOL compiler for DASK.

As a check on the soundness of a flow diagram, it will often be wise to work out at least one of the numerical test examples with numbers which have a reasonable size from a physical point of view.

Example

Illustrating the foregoing ideas and viewpoints, the manner in which we present one of our examples to the student will be described. The problem is how to work out a flow diagram (Fig. 4) for the computation of the dimensions of a high-voltage disk-coil transformer winding, corresponding to the axial-radial cross section in Fig. 2.

The program should fulfill the following conditions: For a given set of insulation data (such as conductor insulation thickness, ducts between coils, and height of the winding), and for a given set of electrical data (such as number of turns, locations of taps, and current density), it should determine the axial and radial dimensions of a disk-coil winding, while taking account of economic and technical considerations (such as a maximum number of six special coils nearest each yoke, winding pressure, and allowable coil width tolerances), as well as considerations regarding losses (such as stray-load losses) and winding temperatures below certain limits.

It should be possible to design the

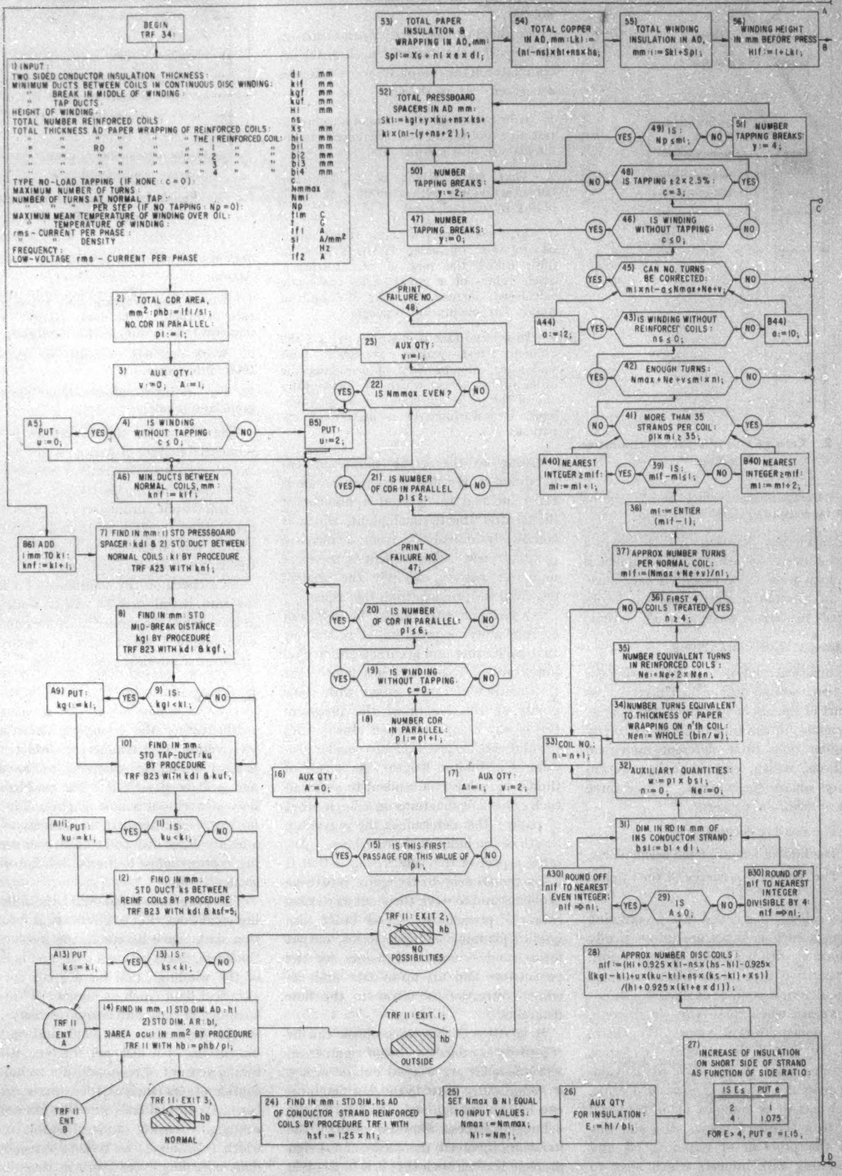

BEGIN TRF 34;
1) INPUT:
TWO SIDED CONDUCTOR INSULATION THICKNESS: di mm
MINIMUM DUCTS BETWEEN COILS IN CONTINUOUS DISC WINDING: kif mm
" BREAK IN MIDDLE OF WINDING: kgf mm
" TAP DUCTS: kuf mm
HEIGHT OF WINDING: Hl mm
TOTAL NUMBER REINFORCED COILS: ns
TOTAL THICKNESS AD PAPER WRAPPING OF REINFORCED COILS: xs mm
" " " " " " THE i REINFORCED COIL: hil mm
" " RD " " " " 1 " " bi1 mm
" " " " " " " 2 " " bi2 mm
" " " " " " " 3 " " bi3 mm
" " " " " " " 4 " " bi4 mm
TYPE NO-LOAD TAPPING (IF NONE: c = 0): c
MAXIMUM NUMBER OF TURNS: Nmmax
NUMBER OF TURNS AT NORMAL TAP: Nml
" " " PER STEP (IF NO TAPPING: Np = 0): Np
MAXIMUM MEAN TEMPERATURE OF WINDING OVER OIL: tlm C
" TEMPERATURE OF WINDING: t C
rms - CURRENT PER PHASE: If1 A
" " DENSITY: sl A/mm²
FREQUENCY: f Hz
LOW-VOLTAGE rms - CURRENT PER PHASE: If2 A
2) TOTAL CDR AREA, mm²: phb := If1/sl; NO. CDR IN PARALLEL: pl := 1;
3) AUX QTY: v := 0; A := 1;
4) IS WINDING WITHOUT TAPPING: c ≤ 0;
A5) PUT: u := 0;
B5) PUT: u := 2;
A6) MIN. DUCTS BETWEEN NORMAL COILS, mm: knf := kif;
B6) ADD 1 mm TO kl: knf := kl + 1;
7) FIND IN mm: 1) STD PRESSBOARD SPACER: kdl & 2) STD DUCT BETWEEN NORMAL COILS: kl BY PROCEDURE TRF A23 WITH knf;
8) FIND IN mm: STD MID-BREAK DISTANCE kgl BY PROCEDURE TRF B23 WITH kdl & kgf;
9) IS: kgl < kl;
A9) PUT: kgl := kl;
10) FIND IN mm: STD TAP-DUCT: ku BY PROCEDURE TRF B23 WITH kdl & kuf;
11) IS: ku < kl;
A11) PUT: ku := kl;
12) FIND IN mm: STD DUCT ks BETWEEN REINF COILS BY PROCEDURE TRF B23 WITH kdl & ksf=5;
13) IS: ks < kl;
A13) PUT ks := kl;
TRF 11 ENT A
14) FIND IN mm, 1) STD DIM. AD: hl 2) STD DIM. AR: bl, 3) AREA ocul IN mm² BY PROCEDURE TRF 11 WITH hb := phb/pl;
TRF 11 ENT B
TRF 11: EXIT 1; hb OUTSIDE
TRF 11: EXIT 2; hb NO POSSIBILITIES
TRF 11: EXIT 3; hb NORMAL
15) IS THIS FIRST PASSAGE FOR THIS VALUE OF pl;
16) AUX QTY: A := 0;
17) AUX QTY: A := 1; v := 2;
18) NUMBER CDR IN PARALLEL: pl := pl + 1;
19) IS WINDING WITHOUT TAPPING: c = 0;
20) IS NUMBER OF CDR IN PARALLEL: pl ≤ 6;
PRINT FAILURE NO. 47;
21) IS NUMBER OF CDR IN PARALLEL pl ≤ 2;
22) IS Nmmax EVEN?
23) AUX QTY: v := 1;
PRINT FAILURE NO. 48;
24) FIND IN mm: STD DIM. hs AD OF CONDUCTOR STRAND REINFORCED COILS BY PROCEDURE TRF 1 WITH hsf := 1.25 × hl;
25) SET Nmax & Nl EQUAL TO INPUT VALUES: Nmax := Nmmax; Nl := Nml;
26) AUX QTY FOR INSULATION: E := hl/bl;
27) INCREASE OF INSULATION ON SHORT SIDE OF STRAND AS FUNCTION OF SIDE RATIO:
IS E ≤ | PUT e
2 | 1
4 | 1.075
FOR E > 4, PUT e = 1.15;
28) APPROX NUMBER DISC COILS: nlf := (Hl + 0.925 × kl − ns × (hs − hl) − 0.925 × ((kgl − kl) + u × (ku − kl) + ns × (ks − kl) + Xs)) / (hl + 0.925 × (kl + e × dl));
29) IS A ≤ 0;
A30) ROUND OFF nlf TO NEAREST EVEN INTEGER: nlf ⇒ nl;
B30) ROUND OFF nlf TO NEAREST INTEGER DIVISIBLE BY 4: nlf ⇒ nl;
31) DIM. IN RD IN mm OF INS CONDUCTOR STRAND: bsl := bl + dl;
32) AUXILIARY QUANTITIES: w := pl × bsl; n := 0; Ne := 0;
33) COIL NO.: n := n + 1;
34) NUMBER TURNS EQUIVALENT TO THICKNESS OF PAPER WRAPPING ON n'th COIL: Nen := WHOLE (bin/w);
35) NUMBER EQUIVALENT TURNS IN REINFORCED COILS: Ne := Ne + 2 × Nen;
36) FIRST 4 COILS TREATED n ≥ 4;
37) APPROX NUMBER TURNS PER NORMAL COIL: mlf := (Nmax + Ne + v)/nl;
38) ml := ENTIER (mlf − 1);
39) IS: mlf − ml ≤ 1;
A40) NEAREST INTEGER ≥ mlf: ml := ml + 1;
B40) NEAREST INTEGER ≥ mlf: ml := ml + 2;
41) MORE THAN 35 STRANDS PER COIL: pl × ml ≥ 35;
42) ENOUGH TURNS: Nmax + Ne + v ≤ ml × nl;
43) IS WINDING WITHOUT REINFORCED COILS: ns ≤ 0;
A44) a := 12;
B44) a := 10;
45) CAN NO. TURNS BE CORRECTED: ml × nl − a ≤ Nmax + Ne + v;
46) IS WINDING WITHOUT TAPPING: c ≤ 0;
47) NUMBER TAPPING BREAKS: y := 0;
48) IS TAPPING ± 2 × 2.5%: c = 3;
49) IS: Np ≤ ml;
50) NUMBER TAPPING BREAKS: y := 2;
51) NUMBER TAPPING BREAKS: y := 4;
52) TOTAL PRESSBOARD SPACERS IN AD mm: Skl := kgl + y × ku + ns × ks + kl × (nl − (y + ns + 2));
53) TOTAL PAPER INSULATION & WRAPPING IN AD, mm: Spl := Xs + nl × e × dl;
54) TOTAL COPPER IN AD, mm: Lkl := (nl − ns) × hl + ns × hs;
55) TOTAL WINDING INSULATION IN AD, mm: i := Skl + Spl;
56) WINDING HEIGHT IN mm BEFORE PRESS Hlf := i + Lkl;
YES
NO
A
B
C
D

Fig. 4. Detailed flow diagram of high-voltage disk coil winding

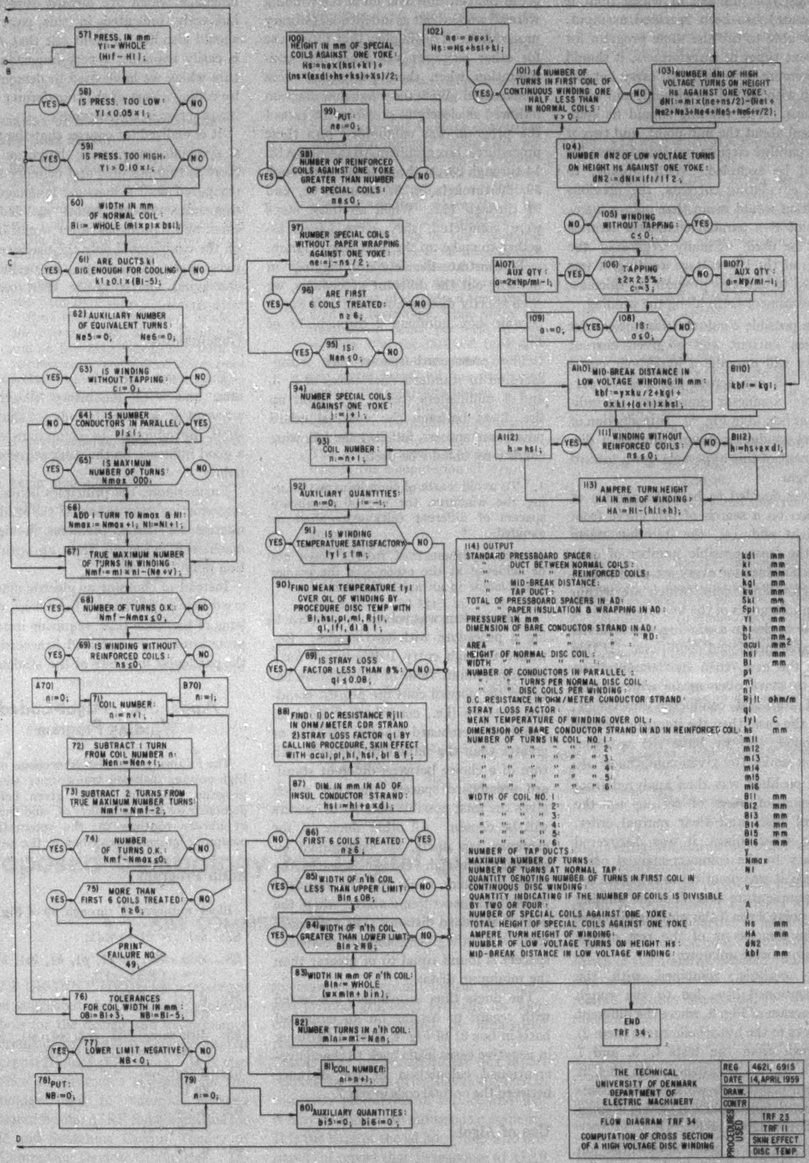

A
B
C
D
57) PRESS. IN mm: Y1:= WHOLE (H1f − H1);
58) IS PRESS. TOO LOW: Y1 < 0.05 × 1;
59) IS PRESS. TOO HIGH: Y1 > 0.10 × 1;
60) WIDTH IN mm OF NORMAL COIL: B1:= WHOLE (m1 × p1 × bs1);
61) ARE DUCTS k1 BIG ENOUGH FOR COOLING: k1 ≥ 0.1 × (B1−5);
62) AUXILIARY NUMBER OF EQUIVALENT TURNS: Ne5:=0; Ne6:=0;
63) IS WINDING WITHOUT TAPPING: c:=0;
64) IS NUMBER CONDUCTORS IN PARALLEL: p1 ≤ 1;
65) IS MAXIMUM NUMBER OF TURNS: Nmax ODD;
66) ADD 1 TURN TO Nmax & N1: Nmax:= Nmax+1; N1:=N1+1;
67) TRUE MAXIMUM NUMBER OF TURNS IN WINDING: Nmf:= m1 × n1 − (Ne+v);
68) NUMBER OF TURNS O.K.: Nmf − Nmax ≤ 0;
69) IS WINDING WITHOUT REINFORCED COILS: ns ≤ 0;
A70) n:=0;
B70) n:=1;
71) COIL NUMBER: n:= n+1;
72) SUBTRACT 1 TURN FROM COIL NUMBER n: Nen:= Nen+1;
73) SUBTRACT 2 TURNS FROM TRUE MAXIMUM NUMBER TURNS: Nmf:= Nmf−2;
74) NUMBER OF TURNS O.K.: Nmf − Nmax ≤ 0;
75) MORE THAN FIRST 6 COILS TREATED: n ≥ 6;
PRINT FAILURE NO. 49;
76) TOLERANCES FOR COIL WIDTH IN mm: OB:= B1+3; NB:= B1−5;
77) LOWER LIMIT NEGATIVE: NB < 0;
78) PUT: NB:=0;
79) n:=0;
80) AUXILIARY QUANTITIES: bi5:=0; bi6:=0;
81) COIL NUMBER: n:=n+1;
82) NUMBER TURNS IN n'th COIL: m1n:= m1 − Nen;
83) WIDTH IN mm OF n'th COIL: B1n:= WHOLE (w × m1n + b1n);
84) WIDTH OF n'th COIL GREATER THAN LOWER LIMIT: B1n ≥ NB;
85) WIDTH OF n'th COIL LESS THAN UPPER LIMIT: B1n ≤ OB;
86) FIRST 6 COILS TREATED: n ≥ 6;
87) DIM. IN mm IN AD OF INSUL CONDUCTOR STRAND: hs1:= h1 + e × d1;
88) FIND: 1) DC RESISTANCE Rj11 IN OHM/METER CDR STRAND 2) STRAY LOSS FACTOR q1 BY CALLING PROCEDURE, SKIN EFFECT WITH acu1, p1, h1, hs1, b1 & f;
89) IS STRAY LOSS FACTOR LESS THAN 8%: q1 ≤ 0.08;
90) FIND MEAN TEMPERATURE ty1 OVER OIL OF WINDING BY PROCEDURE DISC TEMP WITH B1, hs1, p1, m1, Rj11, q1, if1, d1 & f;
91) IS WINDING TEMPERATURE SATISFACTORY: ty1 ≤ tm;
92) AUXILIARY QUANTITIES: n:=0; j:= −1;
93) COIL NUMBER: n:= n+1;
94) NUMBER SPECIAL COILS AGAINST ONE YOKE: j:= j+1;
95) IS: Nen ≤ 0;
96) ARE FIRST 6 COILS TREATED: n ≥ 6;
97) NUMBER SPECIAL COILS WITHOUT PAPER WRAPPING AGAINST ONE YOKE: ne:= j − ns/2;
98) NUMBER OF REINFORCED COILS AGAINST ONE YOKE GREATER THAN NUMBER OF SPECIAL COILS: ne ≤ 0;
99) PUT: ne:=0;
100) HEIGHT IN mm OF SPECIAL COILS AGAINST ONE YOKE: Hs:= ne × (hs1+k1) + (ns × (e × d1 + hs + ks) + Xs)/2;
101) IS NUMBER OF TURNS IN FIRST COIL OF CONTINUOUS WINDING ONE HALF LESS THAN IN NORMAL COILS: v > 0;
102) ne:= ne+1; Hs:= Hs + hs1 + k1;
103) NUMBER dN1 OF HIGH VOLTAGE TURNS ON HEIGHT Hs AGAINST ONE YOKE: dN1:= m1 × (ne + ns/2) − (Ne1 + Ne2 + Ne3 + Ne4 + Ne5 + Ne6 + v/2);
104) NUMBER dN2 OF LOW VOLTAGE TURNS ON HEIGHT Hs AGAINST ONE YOKE: dN2:= dN1 × if1/if2;
105) WINDING WITHOUT TAPPING: c ≤ 0;
106) TAPPING ±2 × 2.5%: c:=3;
A107) AUX QTY: a:= 2 × Np/m1 − 1;
B107) AUX QTY: a:= Np/m1 − 1;
108) IS: a ≤ 0;
109) a:=0;
A110) MID-BREAK DISTANCE IN LOW VOLTAGE WINDING IN mm: kbf:= y × ku/2 + kg1 + a × k1 + (a+1) × hs1;
B110) kbf:= kg1;
111) WINDING WITHOUT REINFORCED COILS: ns ≤ 0;
A112) h:= hs1;
B112) h:= hs + e × d1;
113) AMPERE TURN HEIGHT HA IN mm OF WINDING: HA:= H1 − (h11 + h);
114) OUTPUT
STANDARD PRESSBOARD SPACER: kd1 mm
" DUCT BETWEEN NORMAL COILS: k1 mm
" " " REINFORCED COILS: ks mm
" MID-BREAK DISTANCE: kg1 mm
" TAP DUCT: ku mm
TOTAL OF PRESSBOARD SPACERS IN AD: Sk1 mm
" " PAPER INSULATION & WRAPPING IN AD: Sp1 mm
PRESSURE IN mm Y1 mm
DIMENSION OF BARE CONDUCTOR STRAND IN AD: h1 mm
" " " " " " RD: b1 mm
AREA " " " " : acu1 mm²
HEIGHT OF NORMAL DISC COIL: hs1 mm
WIDTH " " " " : B1 mm
NUMBER OF CONDUCTORS IN PARALLEL: p1
" " TURNS PER NORMAL DISC COIL: m1
" " DISC COILS PER WINDING: n1
D.C. RESISTANCE IN OHM/METER CONDUCTOR STRAND: Rj11 ohm/m
STRAY LOSS FACTOR: q1
MEAN TEMPERATURE OF WINDING OVER OIL: ty1 C
DIMENSION OF BARE CONDUCTOR STRAND IN AD IN REINFORCED COIL: hs mm
NUMBER OF TURNS IN COIL NO. 1: m11
" " " " " " 2: m12
" " " " " " 3: m13
" " " " " " 4: m14
" " " " " " 5: m15
" " " " " " 6: m16
WIDTH OF COIL NO. 1: B11 mm
" " " " 2: B12 mm
" " " " 3: B13 mm
" " " " 4: B14 mm
" " " " 5: B15 mm
" " " " 6: B16 mm
NUMBER OF TAP DUCTS: y
MAXIMUM NUMBER OF TURNS: Nmax
NUMBER OF TURNS AT NORMAL TAP: N1
QUANTITY DENOTING NUMBER OF TURNS IN FIRST COIL IN CONTINUOUS DISC WINDING: v
QUANTITY INDICATING IF THE NUMBER OF COILS IS DIVISIBLE BY TWO OR FOUR: A
NUMBER OF SPECIAL COILS AGAINST ONE YOKE: j
TOTAL HEIGHT OF SPECIAL COILS AGAINST ONE YOKE: Hs mm
AMPERE TURN HEIGHT OF WINDING: HA mm
NUMBER OF LOW VOLTAGE TURNS ON HEIGHT Hs: dN2
MID-BREAK DISTANCE IN LOW VOLTAGE WINDING: kbf mm
END TRF 34;
YES
NO
THE TECHNICAL UNIVERSITY OF DENMARK DEPARTMENT OF ELECTRIC MACHINERY
REG 4621, 6915
DATE 14, APRIL 1959
DRAW.
CONTR
FLOW DIAGRAM TRF 34
COMPUTATION OF CROSS SECTION OF A HIGH VOLTAGE DISC WINDING
PROCEDURES USED
TRF 23
TRF 11
SKIN EFFECT
DISC TEMP.

Fig. 4 (continued)

winding with reinforced coils (as well as without), the number of which (four is maximum) has been specified as input. To be able to use the same program for different types of transformers, it should furthermore be possible to carry out the design with, as well as without, mid-winding taps. The winding should be symmetrical about the mid-point, and two or more parallel circuits per transformer leg should be excluded. With tapping, to avoid circulating currents, the previous condition would mean that only a maximum of two parallel conductor strands could be used. Finally, regarding the design of the low-voltage winding, it was decided to end this program by calculating the ampere-turns along the winding.

The possible conductor dimensions for a given current and a given current density will describe a hyperbolic curve which will cut across the area of permissible conductor dimensions (as indicated at box 14 in Fig. 4). Introduction of parallel conductor strands implies a displacement of this hyperbola down toward the origin.

It was decided to base the process of iteration on a search of this curve from the right end, i.e., the iteration starts with the least possible number of coils and determines for every set of standard conductor dimensions (which are within certain tolerances of the hyperbola) if it is possible to design a winding which fulfills the above-mentioned conditions. If this is not the case, then the computations have to start over again with a new set of standard conductor dimensions; if it is the case, then the iteration process is stopped, and we have the optimum solution under the given conditions.

The problem was then analyzed from the point of view of setting up the different tests and their mutual order. Among other things, it was discovered (contrary to the common manual practice) that it was practical to calculate the mean temperature over oil of the winding at this point, even if the winding diameter and the dimensions of the rest of the transformer were unknown.

This analysis, combined with the aforementioned idea, led to the rough flow diagram of Fig. 3, where the different feedbacks to the hyperbolic curve (item 2) are shown from the tests 4, 5, and 7 through 9. In accordance with this, it was then possible to work out the complete and detailed solution (which is given in the flow diagram of Fig. 4).

From a practical point of view, this job was executed in the following way: the search of the hyperbola, the standardization of the ducts, the calculation of the stray-loss factor, and the calculation of the mean temperature over oil of the winding were all worked out as independent library procedures. This implies that future use for other purposes was taken into consideration when these procedures were constructed. With the rough flow diagram followed as closely as possible, the rest of the program was subdivided into these procedures: box numbers A6 through A13, 14 through 23, 24 through 41, 42 through 59, 60 through 76, 77 through 91, and 92 through 113. When all these steps were completed, they were joined together to make up the final flow diagram.

To illustrate the detailed "digging" in working out the different procedures, we shall shortly discuss how item 1 of Fig. 3 became box numbers 2 through 13 of Fig. 4.

The pressboard spacers are manufactured in standard thicknesses of 2, 3, and 4 millimeters (mm). To build up the ducts between the coils, standard pressboard spacers, fulfilling the following conditions, should be used.

1. To avoid waste of time in manufacturing the windings, the use of pressboard spacers of different thicknesses shall be avoided.
2. Accomplishment of item 1 must not be achieved at the expense of the copper space factor in an axial direction.

The problem was solved by constructing a procedure with two entries. Entrance through entry 1 determined for a given minimum value of the distance between the normal coils the nearest standard distance (i.e., composed of identical standard pressboard spacers) equal to or greater than the minimum distance. In case of a choice between different standard pressboard spacers, the type which gives the least possible number of spacers will be chosen. All other ducts as, for example, tap ducts, are determined by entrance through entry 2 with the given minimum value and the standard pressboard spacer found above, where the nearest standard distance (i.e., a distance composed of the given standard pressboard spacer) is found equal to or greater than the minimum distance.

The ducts thus determined are tested with regard to cooling on an empirical basis in box 61 of Fig. 4, where a feedback in negative cases leads back to the abovementioned calculation of the distance between the normal coils in box 7.

Use of Algol

From the point of view of the user, we are completely convinced of ALGOL's suitability in programing engineering problems. Besides all the probabilities of this language which are already directly or indirectly indicated in this paper, we should also like to point out that ALGOL is easily readable, a very important feature which we have tried to demonstrate by coding a part of the transformer winding program in the Appendix.

It is realized, of course, that the use of a symbolic programing language costs more in running time and memory space, compared with a machine language, but this extra expense seems justified, because many problems never would be put on the computer if only the machine code were used. In addition, the problem of debugging has certainly been reduced with ALGOL.

Conclusions

Experience with thesis work in this area and with laboratory reports in adjacent areas seems to indicate that this short Introduction to computers exerts a good influence on the student's working habits.

Furthermore, the principles behind the Introduction seem to arouse the student's interest for design problems in electric machinery, especially from an optimization point of view.

Instead of choosing an electric machinery design problem as we have done, problems of the same type, requiring iterative procedures, may be selected, in accord with the particular interests of the students.

Appendix. Algol-Coded Winding Program

The following are selected sections of the high-voltage disk-coil transformer winding program coded in ALGOL, given here to demonstrate the readability of, and the use of library routines in, the international computer language described in the paper.

Main Program

Block numbers 87 through 99 of Fig. 4.

```
87: hs1 := h1+e×d1;
88: skin effect (acu1, p1, h1, hs1, b1, f)
        =: (Rj11, q1);
89: if (q1 > 0.08);
    go to B14;
90: disk temp (B1, hs1, p1, m1, Rj11,
        q1, If1, d1, t) =: (ty1);
91: if (ty1>tm);
    go to B14;
92: n := 0;
    j := −1;
93: n := n+1;
94: j := j+1;
95: if (Nen≤0);
    go to 97;
```

```
96:  if (n<6);
     go to 93;
97:  ne := j-ns/2;
98:  if (ne>0);
     go to 100;
99:  ne := 0;
```

Calculation of D-C Resistance and Stray-Loss Factor

```
     procedure skin effect (acun, pn, hn,
     hsn, bn, f) =: (Rjn1, qn);

     comment acun is the area in square
     mm and hn and bn the height and
     width in mm of the bare conductor,
     while hsn is the height in mm of
     the insulated conductor.  pn is the
     number of conductors in parallel
     and f the frequency.
       Rjn1 is the d-c resistance in ohms
     per meter length of the conductor
     and qn the stray-loss factor, both at
     the reference temperature 75 degrees
     centigrade.  v is the reduced height
     of the conductor at the same tempera-
     ture.
       Reference: L. Hyldgaard-Jensen:
     Transformatoren, Appendix I, Copen-
     hagen 1954;

     begin skin effect:
      Rjnl := 2.17₁₀-2/acun;
     a(X) := X×(exp (2×X)-exp
             (-2×X)+2×sin (2×
             X))/(exp (2×X)+exp
             (-2×X)-2×cos (2×X));
     b(Y) := 2×Y×(exp (Y)-exp
             (-Y)-2×sin (Y))/(exp
             (Y)+exp (-Y)+2×cos
             (Y));
       cs := 95.5₁₀-3×sqrt(f×hn/(50×
             hsn));
        v := bn×cs;
       qn := -1+a(v)+(pn↑2↓-1)×
             b(v)/3;
     return end skin effect;
```

Temperature of Winding Over Oil

```
     procedure layer temp (Bn, hsn, pn,
     T2, Rjn1, qn, Ifn, t) =: (tyn),
     disk temp (Bn, hsn, pn, mn, Rjn1,
     qn, Ifn, t) =: (tyn);
     comment The mean temperature tyn
     over oil of the winding is computed
     by Gotter's method (G. Gotter:
     Erwarmung und Kuhlung elektrischen
     Maschinen, Springer Verlag 1954, pp.
     198-204) by considering a 1-meter-
     long section of a thread in a layer-
     winding respectively of a coil in a
     disk-coil winding, where Bn is the
     width and hsn the height of the
     section, while dn is the conductor
     insulation thickness, all three in mm.
       pn is the number of conductors
     parallel in radial direction.  T2 is
     the number of threads, if the winding
     is of the layer-type, while mn is the
     number of turns per coil, if the
     winding is of the disk-coil type.
       Rjn1 is the d-c resistance in ohms
     per meter length of the conductor
     and qn the stray-loss factor.  Ifn is
     the current in amperes per phase,
     and t is the maximum temperature of
     the winding in degrees centigrade.
     The covered area has been put to
     0.30 in the computation of Afn;
     begin layer temp:
     mn := 1;
     I := Ifn/T2;
     go to 75;
     begin disk temp:
     I := Ifn;
 75: if (t≤75);
     begin a0 := 70.724; a1 := 17.165;
       a₂ := -3.2939; a3 := 0.27769;
     go to RD end;
105: if (t ≤ 105);
     begin a0 := 79.570; a1 := 22.725;
       a2 := -3.9282; a3 := 0.31699;
     go to RD end;
     stop;
 RD: RD(X) := ((a3×X+a2)×X+a1)
       ×X+a0;
     comment RD(X) is the heat con-
     vection factor in radial direction as
     a function of the loss in watts per
     square meter surface area of the
     section (Gotter: Abb. 94);
         b0 := 1.0022; b1 := -2.4544₁₀
         -2; b2 := 3.4928₁₀-4; b3 :=
         -2.3164₁₀-6; b4 := 5.8695₁₀
         -9;
 AD: AD(Y) := (((b4×Y+b3)×Y+b2)
       ×Y+bl)×Y+b0;
     comment AD(Y) is the heat con-
     vection factor in axial direction
     divided by the heat convection factor
     in radial direction as a function of
     the width Bn of the section (Gotter:
     Abb. 95);
     Afn  := 1.4₁₀-3×(Bn+hsn);
     W    := mn×Rjn1×(1+qn)×
             I↑2↓/pn;
     w    := W/Afn;
     myr  := RD(w);
     f    := AD(Bn);
     mym  := myr×(1+f×Bn/hsn)/
             (1+Bn/hsn);
     if (dn≤1);
     begin tyn := w/mym;
     return end layer and disk temp;
     tyn  := w×(2.75₁₀-3×(dn-1)+
             1/mym;
     return end layer and disc temp;
```

References

1. Preliminary Report—International Algebraic Language, A. J. Perlis, K. Samelson. *Communications*, Association for Computing Machinery, New York, N. Y., vol. 1, no. 12, Dec. 1958.

2. Philosophy of Applying Digital Computers to the Design of Electrical Apparatus, P. A. Abetti, W. J. Cuthbertson, S. B. Williams. *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 77, July 1958, pp. 367–79.

3. The Practice and the Economics of Applying Digital Computers to Engineering Problems, P. A. Abetti, S. B. Williams. *Ibid.*, pp. 331–42.

4. Digital Computer Programming (book), D. D. McCracken. John Wiley & Sons, Inc., New York, N. Y., 1957.

Proposed Symbology for Digital Systems

AIEE COMMITTEE REPORT

THE URGENT NEED for a reference on suggested algebraic and graphical symbols for the design and maintenance requirements of digital systems prompted the preparation of this report. It is also intended to serve as a reference for the symbology used in tutorial lectures sponsored by the AIEE Subcommittee for Logic and Switching Circuit Theory of the Computing Devices Committee. Included are algebraic symbols for representing variables and connectives used in logical expressions; graphical symbols for manually prepared logical diagrams; and printed graphical symbols for logical diagrams prepared by high-speed printers.

It is fully appreciated that this report reflects the opinon of engineers engaged in computer production or service and that it may be subject to numerous comments. It is the hope of the ad hoc group who prepared the report that, in response to AIEE policy of inviting written comment, it will provoke discussion and thus serve as a direct contribution to the formulation of engineering standards by other authorized organizations.

In preparing the report the philosophy of U. S. Government agencies and the U. S. Air Force in particular as the largest users, has been taken into consideration. Much of the material is based on the findings of a study by the Air Force,[1] and on related military standards, for example, reference 2.

Paper 60-1224, recommended by the AIEE Computing Devices Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Chicago, Ill., October 9–14, 1960. Manuscript submitted June 3, 1960; made available for printing September 23, 1960.

The members of the Ad Hoc Working Group of the Logic and Switching Circuit Theory Subcommittee of the AIEE Computing Devices Committee are T. H. Mott, Jr., *chairman*; T. C. Bartee, D. Drusdow, R. J. Preiss, E. J. Schubert, W. J. Smith, and T. W. Sze.

Table I. Symbols for Logical Operations

Operation	Symbol	Example
OR	$+$	$A+B$ (A OR B)
AND	$\cdot$	$A \cdot B$ (A AND B)
NOT	$'$	A' (NOT A)
EQUIVALENCE	$=$	$A+A=A$

Ground Rules for Choice of Symbols

Three major divisions are proposed as follows, in which the choice of symbols is based on the following rules:

1. *Algebraic Symbols:* Algebraic symbols shall designate the relations of logical variables in binary notation for combinational and sequential logic. Appropriate indexing for canonical forms (minterms and maxterms), general logical expressions, and time-dependent expressions is tentatively provided. All symbols are restricted to those types which are available on conventional typewriters and printers' keyboards.
2. *Graphical Symbols:* Drafting symbols are suggested for manually prepared logical diagrams with provision for complete terminal, signal, and functional identification. The shape of the symbols is restricted to those formed by combinations of horizontal and vertical lines, and circles.
3. *Printed Graphical Symbols:* Printed graphical symbols are equivalent to the drafting symbols in the next section but are prepared by programed printing devices. The shape of the symbols is dictated by the restriction to the minimum number of additional characters.

Algebraic Symbols for Logic Equations

The proposed algebraic symbols, used to designate relations between logical variables, and restricted to those available on conventional typewriters and printers' keyboards, are divided into four categories: logical variables, logical operations, logical expressions, and time-dependent expressions.

Logical Variables

As in conventional algebra, the logical variables are represented by letter symbols with or without subscripts. Capital or lower-case letters may be used, e.g., A, B, C, a, b, c, etc. However, when both logical and algebraic variables are present in a text, it is recommended that the capital letters be assigned to the logical variables and the lower-case letters to the algebraic variables. For example:

(Logical variables)		(Algebraic variables)
$T=A+B$	when	$b^2-4ac \geq 0$
$T=A \cdot B$	when	$b^2-4ac<0$

Associated with logical variables are the values which they take on. In binary logic, these values are represented by 1 and 0 (truth and falsity, active and inactive, etc). The Greek letter ϕ should be used if the designer "doesn't care" whether the value of a variable is 1 or 0.

Logical Operations

When a logical operation word appears in a text, it should be spelled out in full with capital letters, e. g., A AND B.

For expressing the following logical operations in symbolic form, it is proposed that the symbols listed in Table I be used.

$A+B$ is called the sum of A and B; $A \cdot B$, the product of A and B; and A', the negation of A.

Other operations, such as OR-NOT (NOR), AND-NOT, and EXCLUSIVE OR, are considered as derived operations, and will use the combinations of symbols listed in Table I.

Logical Expressions

General form. General logical expressions are built up from logical operations, logical variables, parentheses, and brackets. Where parentheses and brackets are omitted in an expression, the hierarchy of operations is assumed to be: $'$, $\cdot$, $+$, $=$.

Logical functions are associated with logical expressions as follows: Suppose A, B, C, D, . . . are the variables appearing in a logical expression; then, the expression specifies the value of the function for the arguments A, B, C, D, ... For instance, suppose $A \cdot B' + C \cdot D$ is a logical expression; then, for the different values of the arguments A, B, C, and D, the values $f(A, B, C, D)$ of the corresponding function f are calculated from the expression $A \cdot B' + C \cdot D$.

Canonical form. Associated with any given logical function are many different logical expressions. Assume the arguments of the function expressed by a fixed assignment of variables; then, two of the logical expressions associated with each function are defined as canonical. The "sum of products" canonical form is the sum of certain products called minterms. These are products in which all of the assigned variables appear exactly once, negated or unnegated. The dual canonical form is defined analogously as a "product of sums."

Positional notation. The minterms of a function correspond in obvious fashion to those value assignments of the arguments which give the function the value 1 (or truth). Hence, instead of expressing the function in canonical form as a sum of products it is sometimes convenient to represent it in terms of a set of (0, 1) n-tuples (where n is the number of arguments of the function). Analogously, introducing a third symbol,—(dash) or 2 or x, to indicate that the corresponding argument (variable) is not present in the product, enables one to represent products in general logical expressions as (0, 1,—) n-tuples.

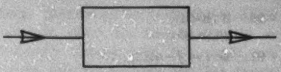


Fig. 1. Illustrating use of arrowheads for direction of signal flow; not to be placed immediately adjacent to any graphical symbol, input, or output

For instance, the 3-argument function expressed by $A' \cdot B' \cdot C' + B \cdot C$ has a canonical form $A' \cdot B' \cdot C' + A \cdot B \cdot C + A' \cdot B \cdot C$. Hence the function may be denoted by the set of triples 000, −11, or alternately by the set 000, 111, 011.

Time-Dependent Expressions

It is proposed that superscripts be reserved to denote values of time variables. For example, $f^{t+1}(A, B, C) = (A \cdot B + C)^t$ or $= A^t \cdot B^t + C^t$.

Graphical Symbols for Logic Diagrams

This section contains a system of drafting symbols proposed for digital circuits. Each symbol represents a "binary logical function," and is a graphical representation of an electrical function. The list presents basic logic symbols in their simplest form, and is consistent with other accepted standards and conventions. The system of symbols is fundamental in nature and therefore has growth potential. The basic symbols were selected with two objectives in mind:

1. Specific (geometrical) shape.
2. "Logical" associations to functional circuits.

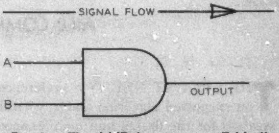


Fig. 2. The AND function; see Table II

Table II. The AND Function of Fig. 2

Input		Output
A	B	$X=A \cdot B$
0	0	0
0	1	0
1	0	0
1	1	1

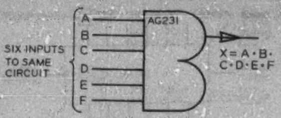


Fig. 3. Multiple input AND gate

All graphical symbols contained in this section are drawn to the same scale, and rectangular symbols are in specific aspect ratio.

State of Signal and Direction of Signal Flow

A signal may assume either the "active or true" state (1) or the "inactive or false" state (0). The electric signal levels (high or low) for the true or false state must be specified explicity for the individual circuit diagrams.

1. The direction of signal flow shall be indicated by arrowheads superimposed on the line channel path in all cases where direction of flow is not obvious.
2. When the direction of signal flow is obvious from the associated logic, the arrowheads may be omitted.
3. Arrowheads shall not be placed immediately adjacent to any graphical symbol, input or output; see Fig. 1.

Basic Logic Symbol Reference States

In this report all basic logic symbols shall be considered to operate for the input signals being true (1), high level. The true (1) high-level state of the input signals activates the logic element in the manner indicated by the specific symbol. This convention shall be followed regardless of the actual mode of operation of the physical devices used to implement the logic. Small circles adjacent to the symbols shall indicate inversion of the signal, i.e., low level.

The presence of a small circle at the input of the logic element indicates that it will be activated by the (0) low-level signal. Absence of the small circle at the input indicates that (1) high-level is the activating signal. Accordingly, a small circle at the output indicates that the function is the inverse of the normal symbol function.

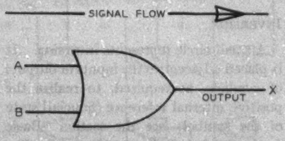


Fig. 4. The OR function; see Table III

Table III. The OR Function of Fig. 4

Input A	Input B	Output X = A + B
0	0	0
0	1	1
1	0	1
1	1	1

Graphical Symbol Shape and Size Considerations

The size and shape of each graphical symbol in this report are determined by the following requirements:

Each symbol must be instantly recognizable in a complex diagram by virtue of a distinctively different geometric shape dictated by its different peripheral configuration.

Each symbol must be evacuous, thus permitting 0.140-inch high-internal letter tagging to identify the following:

1. Logic reference designation; two alphabetical symbols and three digits.
2. Six-digit number which locates circuit by cabinet number, rack, row, and nest.
3. Four-digit number identifying circuit schematic.

Basic Logic Symbols

AND *Gates—Function*

The AND gate is a circuit or device performing the logical AND function (i.e., the Boolean operation of intersection), if and only if all the inputs are in the 1 state; see Fig. 2 and Table II.

Note that:

1. AND gates are indicated by a horizontally elongated letter D, with data flowing from the straight base line to the curved side.
2. The input lines contact the vertical base line at a perpendicular angle.
3. The output line departs from the midpoint of the curved side.
4. Two or more input lines may be shown.
5. Fig. 3 illustrates an application where many input lines are required.
6. The output equals 1 (logical 1) if and only if all the inputs are 1; see Table II.

OR *Gates—Function*

The OR gate is a circuit or device whose output is energized when one or more of the inputs is in its prescribed state. An OR circuit performs the function of the logical OR, which is, in Boolean algebra the operation of union; see Fig. 4 and Table III.

Note that:

1. OR circuits are indicated by an arrowhead type of symbol with data flowing from the curved portion to the point.

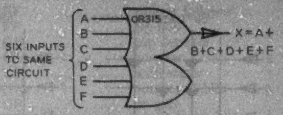


Fig. 5. Multiple input or gate

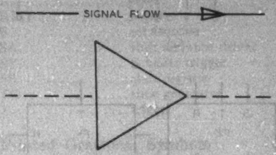


Fig. 6. Amplifier symbol

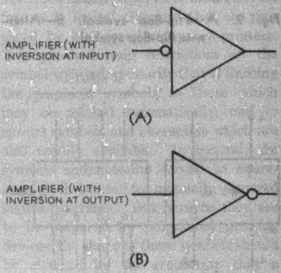

AMPLIFIER (WITH INVERSION AT INPUT)
(A)
AMPLIFIER (WITH INVERSION AT OUTPUT)
(B)

Fig. 7. Amplifier. A—With inversion at input. B—With inversion at output

2. The input lines contact the symbol's curved portion.
3. The output line departs from the point of the symbol.
4. Two or more inputs may be shown.
5. Fig. 5 illustrates an application where many input lines are required.

Amplifier—Function

The amplifier is a device whose output is some function of a single input. Generally the output is an improved version of one or more of the essential features of the input (such as level, signal swing, or rise time). The symbol is shown in Fig. 6.

Note that:

1. The amplifier is shown as an equilateral triangle, pointed in the direction of signal-data flow; see Figs. 6 and 7.
2. The exact function of the device may be indicated by markings within the triangle.
3. Function markings and abbreviations (other than those included within this symbolic system) shall be identified by notation for each drawing.

Examples are shown in Figs. 7(A) and 7(B).

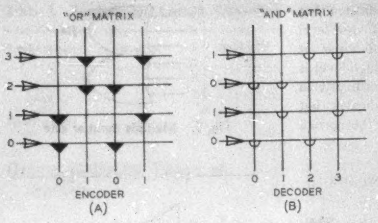


Fig. 8 (left). OR and AND matrices. Symbols for logical network with elements connected at some of its intersections of input-output leads

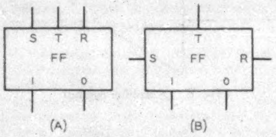


Fig. 9. A—Flip-flop symbol. B—Alternate flip-flop symbol

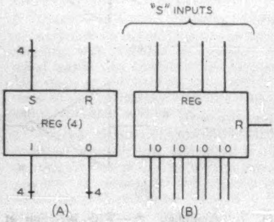


Fig. 10. A—Binary register symbol. B—Alternate binary register symbol

AND/OR *Matrices–Function*

AND/OR matrices are any logical network whose configuration is a rectangular array of intersections of its input-output leads, with elements connected at some of these intersections. The symbols are shown in Fig. 8.

Note that inputs and outputs may be oriented as is convenient.

Storage Circuits—Function

A storage device is able to retain a signal. Conventional storage elements are: flip-flops (single-bit storage), binary registers (*N*-bit storage); and shift registers (a binary register with internal shift provision).

Example 1—Flip-Flop: The flip-flop, Figs. 9(A) and (B), is a bistable circuit capable of storing a single bit of information. It has three possible inputs; set, reset, and trigger; with two possible outputs, 1 and 0.

The two outputs are always of opposite polarity. A 1 is stored in the flip-flop when the 1 output is at the 1 level and the 0 output is at the 0 level. A zero is stored in the flip-fiop when the above condition is reversed.

The flip-flop assumes the 1 state (1 output at conventional 1 level) when an effective signal appears at the S input. It assumes the 0 state when an effective signal appears at the R input. It reverses its state when an effective signal appears at the T input.

Note that:

1. Symbols for trigger circuits are indicated by rectangles which have an aspect ratio of 0.8 to 1.4.
2. The trigger input contacts the center of one broad side, and may be identified by the letter T.
3. The set and reset inputs must be identified by letters S and R respectively:
   S input shall be in proximity to the 1.
   R input shall be in proximity to the 0.
4. The 1 and 0 output lines contact the broad side which is not contacted by an input.

Example 2—Binary Register: This device is a group of flip-flops used in parallel to constitute a single register (as to store four bits of a character).

Note that:

1. The number of bits or individual flip-flops in the register is indicated within parentheses following the abbreviation REG.
2. Fig. 10(A) shows four S inputs grouped on one multiple input line and four each 1 and 0 grouped output lines.
3. In some applications, individual input and output lines are shown as in Fig. 10(B).
4. The aspect ratio of the symbol shall be 1 to 2.5 or greater.

Example 3—Shift Register: This device is a group of flip-flops with provisions to displace or shift the content of the register one stage at a time to the right or left by means of the "shift" inputs.

Note that:

1. The location of shift input at the left corner of the symbol indicates that the shift is from left to right.
2. If the shift is from right to left, the shift input is located at the right corner of the symbol.
3. The aspect ratio shall be to 1 to 2.5 or greater as required.

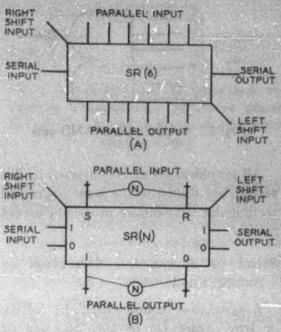


Fig. 11. A—Shift register symbol. B—Alternate shift register symbol

Fig. 12 (left). Inversion symbol

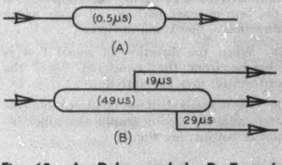


Fig. 13. A—Delay symbol. B—Tapped-delay symbol

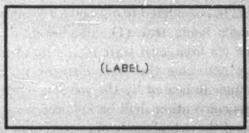


Fig. 14. General logic element symbol

4. The choice of either the symbol in Fig. 11(A) or that in 11(B) is dependent on diagram arrangement of associated symbology.

INVERSION

A small circle represents inversion. It is placed adjacent to the inputs or outputs of symbols, as required, to realize the positive internal reference (fiducial) state of the symbol; see the section "Basic Logic Symbol Reference States." The symbol is shown in Fig. 12.

Delay

Delay is expressed by any passive device which effects a time delay of the signal. The symbol is shown in Fig. 13.

Note that:

1. Delay time indicated within parentheses; see Fig. 13(A).

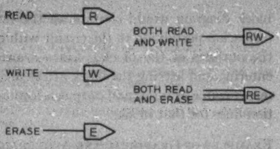

READ
R
BOTH READ AND WRITE
RW
WRITE
W
BOTH READ AND ERASE
RE
ERASE
E

Fig. 15. Magnetic head graphic symbols

2. Fig. 13(B) illustrates a tapped delay line.

3. Aspect ratio generally 5 to 11.

GENERAL LOGIC ELEMENT—FUNCTION

An attempt to assign a logical symbol to the many digital operations and junctions designed into digital devices would cause this system of symbols to become unmanageable and would defy the average person's memory for correct application. Accordingly, a "general logic symbol" with inscribed notations serves as an excellent catchall for any function not classified herein. The symbol is shown in Fig. 14.

Note that:

1. The symbol shall be adequately labeled to describe completely the function performed.

2. The aspect ratio is 1.1 to 2.2.

MAGNETIC HEADS—FUNCTION

Magnetic heads are electromagnetic devices used for magnetizing (recording), detecting (reading), or demagnetizing (erasing) a magnetic surface. The graphic symbols are shown in Fig. 15.

Note that:

1. Orientation of magnetic heads may be in any direction, as convenient.

2. Inscribed capital letters identify functional use of each magnetic head as shown.

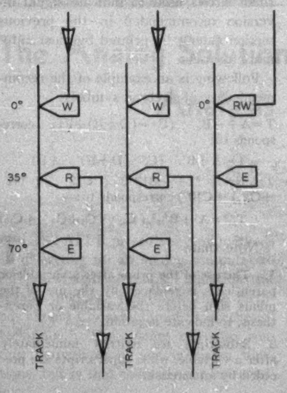

0°
W
W
0°
RW
35°
R
R
E
70°
E
E
TRACK
TRACK
TRACK

Fig. 16 (left). Magnetic drum graphic symbol

MAGNETIC DRUMS—FUNCTION

A rotating cylinder on which information may be stored as magnetically polarized areas, usually along several parallel tracks around the periphery. The graphic symbols are shown in Fig. 16.

Note that:

1. Each track of a magnetic drum, Fig. 16 (or similar recording device), is indicated by a single line terminating in an open arrowhead indicating direction or travel.

2. The angular position of each head is indicated in degrees, or delay time, near the head.

3. The position of all heads are related to the recording head W unless otherwise noted.

EXPANDED LIST OF BASIC LOGIC SYMBOLS

Employing the logic symbols developed so far, Fig. 17 represents an expansion of the system to demonstrate its growth potential.

ABBREVIATIONS

Nondistinctive logic symbols shall be tagged with mnemonic abbreviations (in alphabetical order) to describe the function performed. In the list below the function follows the abbreviation:

AG	AND gate
CT (*N*)	binary counter (*N* bits)
RG (*N*)	register (*N*) stage
BO	blocking oscillator
CF	cathode follower
CD	current driver
DL	delay
EF	emitter follower
FF	flip-flop
GD	gap detector
IV	inverter
LS	level setter
(N)	number of bits
OR	OR gate
OS	one shot
PA	pulse amplifier
PG	pulse generator
R	reset
REG	register
S	set
SA	shift advance driver
ST	Schmitt trigger
SR	shift register
TG	time generator
T	trigger

Printed Graphical Symbols

The graphical symbols and notation described in this paper have been selected for use in logical diagrams that are prepared by high-speed computer printers. This places certain restrictions on the symbology used: primarily that of limiting the graphical symbols to those which may be printed automatically, and to special symbols and characters which are also readily available. In general, the symbols and notation should be equivalent to those used for manually prepared logical diagrams. But because they are to be prepared by programed printing devices, the shape of these symbols should be dictated by the restriction that a minimum number of additional characters need to be added to the most commonly used mechanical or electronic printing devices.

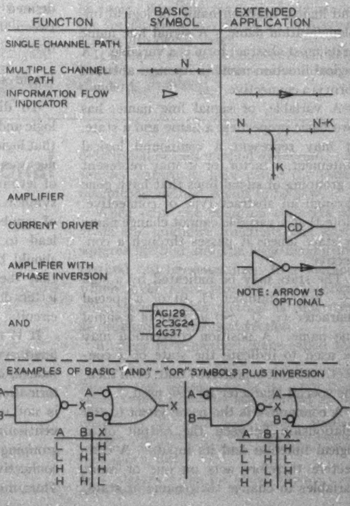

FUNCTION
BASIC SYMBOL
EXTENDED APPLICATION
SINGLE CHANNEL PATH
MULTIPLE CHANNEL PATH
N
INFORMATION FLOW INDICATOR
N
N-K
K
AMPLIFIER
CURRENT DRIVER
CD
AMPLIFIER WITH PHASE INVERSION
NOTE: ARROW IS OPTIONAL
AND
AG129
2C3G24
4G37
EXAMPLES OF BASIC "AND" - "OR" SYMBOLS PLUS INVERSION
A
B
X
A
B
X
A B X
L L H
L H H
H L H
H H L
A
B
X
A
B
X
A B X
L L H
L H L
H L H
H H H

Fig. 17 (right). Expansion of logic symbol system to demonstrate its growth potential

This restriction, then, rules out the use of the shapes of the symbols on the diagrams as a possible means for identifying their functions. Also, since machine-produced records are usually used for more than one purpose, the difference between the pure logic and the other information, such as implementation or placement information, must be clearly kept separate.

With these two aims in mind, the proposed philosophy offered is recommended for solicited AIEE papers:

FLOW OF INFORMATION

At least one major direction of flow of information must be maintained on a logic diagram. If no arrows are shown, information is to be read from left to right, top to bottom.

ALPHABETIC AND NUMERIC INFORMATION

The burden of distinguishing a letter from a number shall always be placed on the letter. Three instances are cited with recommendations: The letter O and the number 0 should be clearly distinguishable on a printed diagram. Similarly, the capital letter I and the number 1, and the letter Z and the number 2 should be clearly distinguishable. It is recommended that the letter O be slashed: Ø; the capital letter I have bars at top and base; and the letter Z appear with a bar through it: Ƶ.

SIGNAL NAME AND FUNCTION NAME

It is recommended that a distinction be maintained between signal line name and logical function name, or logical implementation name. A signal line name in its most abstract form is a variable. A logical function name in its most abstract form is a connective.

A variable, or signal line name, has two basic properties: a name and a state. It may represent a compound logical statement, a factor, or it may represent a grouping of signal lines that have gone through an abstract type of connective. Note that a variable cannot change name or state unless it passes through a connective.

The state may be indicated by a sign (+, −), a number (1, 0) or a special character (*s*, ◊, ~) as part of a signal line name. A position convention may be used to indicate the state of a connective, but it is preferred that, even then, the special character be also used.

A connective is the name given to the relationship between the output of a logical function and its inputs. A connective therefore acts on one or more variables to change their name or state, or both. A connective may exist in three forms: It may be an algebraic connective; it may be a physical implementation of an algebraic, or a group of algebraic connectives; or it may be an abstract, or undefined connective. The difference between these three types must be clearly discernible on the diagram.

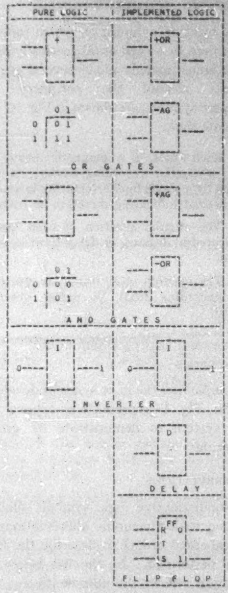


Fig. 18. Machine-produced block diagrams

LOGIC AND CIRCUIT INFORMATION

To distinguish between implemented logic and abstract logic, it is recommended that logical symbols be used as logical connectives, and that letters or combinations of letters be used for implemented logic symbols. The appearance of both types of symbols on one diagram should not lead to confusion. A concerted effort should be made, however, to guarantee that the same letter or combination of letters does not stand for different types of circuits.

It is recommended that a clearly defined symbol should exist between line names, connective names, and other information on the diagram. If a symbol is not used, then a clearly defined convention may be substituted so that the grouping belonging to the name or to the connective be clearly distinguishable. Thus, for instance, a machine-produced block diagram would contain all the information pertaining to the circuit within the outlines of the block. Lines drawn entering and leaving the block are clearly recognizable as input and output information lines for that block.

EXAMPLES OF GRAPHICAL AND ALGEBRAIC SYMBOLS

The types of the graphical symbols used are limited to rectangles by the fact that these symbols are to be automatically printed. Therefore it is necessary to indicate the function yielded by a given block by indicating the function within the rectangle. Fig. 18 illustrates the recommended block diagrams for AND gate, OR gate, FLIP-FLOP, DELAY, and INVERTER.

The symbols shown in Fig. 18 are printed on an automatic printer. Space is left within each rectangle in the implemented logic block so that part number, location in the computer, voltage levels, and other details may also be printed inside the rectangle. The designation for external pins is placed outside the rectangle.

Two types of symbols are shown: 1. Pure logic symbols, for use in diagrams, when only the logic of the machine is to be printed. 2. Implemented logic symbols for diagrams in connection with the construction and maintenance of the machine. For this reason, both positive and negative AND and OR gates are shown in the implemented logic section of the figure. (A negative AND gate corresponds physically to a positive OR gate, and a negative OR gate corresponds physically to a positive AND gate.)

Also, NOR is used to designate an OR gate with an inverted output; and NAG an AND gate with an inverted output. This distinction is necessary because the small circles used to indicate signal inversion recommended in the previous section cannot be printed by most automatic printers.

Following is an example of the recommended use of algebraic symbols:

T=A+−B. (C·−(D+E)+F) corresponds to:

$$T=A+B'.\ (C.\ \ (D+E)'\ \ +F)$$

T11*3=A *1+−B13*2. (C12 .−(G1 +G2) +C10) corresponds to:

$$T_{11}{}^{3}=A^{1}+B'_{12}{}^{2}.(C_{12}.(G_1+G_2)'+C_{10})$$

Note that:

1. The use of the prime after a variable or parenthesis is replaced by the use of the minus sign before the variable or parenthesis, to indicate negation.
2. Subscripts are written immediately after a variable, while superscripts are preceded by an asterisk.

Vol.
Pt. 1
1960

References

1. The Language and Symbology of Digital Computer Systems. Radio Corporation of America Service Company, Camden, N. J., U.S. Air Force Contract no. *AF(604)-1863*, 1959.
2. Electrical and Electronic Symbols. *USAF MIL-STD-15A*, United States Air Force, Washington, D. C.

◆

Discussion

Ralph J. Preiss (International Business Machines Corporation, Poughkeepsie, N. Y.): First, I would suggest that the following definition of a symbol be inserted at the start of the section "Graphical Symbols for Logic Diagrams":

> A symbol is considered as the aggregate of all its parts. It consists of an evacuous shape and an alphanumeric code inside this shape. This alphanumeric code may be used in conjunction with its respective shape, or with the general logic element shape, at the discretion of the user, and consequently is alone capable of fulfilling the requirements and intentions of the standard.

Without this definition, the section "Printed Graphical Symbols" is inconsistent.

Second, I should like to submit a revised list of alphanumeric codes for the abbreviations in the section "Abbreviations," as follows:

A.....and
AM...amplifier
BO....blocking oscillator
C......converter
CT....counter
E.....extender
EF....emitter follower
FF....flip-flop
G.....gate
GD...gap detector
I.....inverter
L.....level setter
O.....or
OE...exclusive or
OS...oscillator (time generator)
PG...pulse generator
RG...register
SR.....shift register
SS.....single shot (one shot)
TD....time delay

Third, all the illustrations representing the symbols should include their alphanumeric codes in accordance with the first two suggestions.

J. F. Kalbach (Burroughs Corporation, Pasadena, Calif.): The symbols proposed in this paper might have been useful to the company with which I am affiliated had they been available when needed. However, in the absence of national and internal company standards, sets of logic symbols have been generated and used by each major engineering activity. Because each activity deals with different types of logic and different digital applications, there are differences between the standards of several engineering divisions within our company.

The symbols used by our particular engineering division are different from the proposed standards. Perhaps we are prejudiced by familiarity, but in comparison with our own, the proposed symbols appear awkward and too large. Drawing space requirements, using the proposed template would greatly exceed the requirements using our own. Perhaps this could be overcome by changing the dimensions of the proposed template. Engineers who are familiar with our own symbology are reluctant to use another.

However, the above considerations become academic to any group which is converting to machine-produced logic block diagrams. We have embarked on such a program and have encountered practically no objections from those who use the rectangular diagrams produced by machine. A program for generating logic block diagrams with variable size and position was described in a paper by Carroll Slemaker in 1960. The logical symbols are all rectangular and are of varying size and position on the page. They have no dots or arrowheads to identify function. The sole identification consists of symbols printed within the rectangular blocks.

In view of the predicted general adoption of machine-printed logic diagrams throughout the industry, we feel that standardization is needed in sets of symbols to be used for identifying rectangular blocks.

G. L. Hollander (Hughes Aircraft Company, Fullerton, Calif.): The ad hoc group should be congratulated on their pioneering action of proposing standard symbology and providing a forum for comments from diverse groups to speed the time of a universally acceptable standard. The selection of specific graphical and algebraic symbols may be controversial, but in time everyone will benefit from the adoption of standards that are unambiguous, consistent, and easily distinguishable in poor drawings and when greatly reduced. However, to be universal, a standard should provide for all present and future approaches to logical design.

The two prevalent approaches to logical design differ by the primary reliance on drawings (pictorial symbology) or equations (notational symbology). Historically the pictorial or the notational approach was favored by the "Eastern" and "Western" factions respectively; automated design procedures also depend on the notational approach.

The proposed symbology is incompatible with the notational approach. The letter symbols for the gating structures do not distinguish between the basic and the inverted functions, such as OR and NOR. Reference to the drawing is essential to differentiate the two. Since, during the last few years, the NOR function has been widely adopted, the standard should provide for a basic literal NOR and NAND designator.

One should also question whether the "basic" operations, described in the section "Logical Operations," are really fundamental or merely historical. If, e.g., the NOR and NAND circuitry become the standard building blocks, designers may gradually consider the negated functions to be basic. Limiting the basic functions in the specifications may freeze an antiquated system into logical design.

It is hoped that these comments will help the ad hoc group establish a more universal standard.

The Analog Solution of Some Functional Analysis Problems

K. K. GRAUPE
NONMEMBER AIEE

THIS PAPER presents several methods of setting up an analog computer to solve a system of coupled equations for several unknown functions; these methods extend Clymer's "implicit synthesis" technique.[1] The system can also contain derivatives of the unknown functions, but in that case only one independent variable is considered. The system of equations need not be linear. Only the existence of certain derivatives of the given functions is assumed.

The principal method, here called the least-squares method, is mentioned in an abstract,[2] but was evidently applied only to linear systems. No practicable description of that method was found. A simpler variation, here called the least-magnitude method, seems to be new.

When only one equation connects all given and unknown functions, the least-squares method and the least-magnitude method can be generalized to a method which gives more freedom in the computer setup. This method is described as problem 1 in the next section.

Another new method, here called the

Paper 60-1230, recommended by the AIEE Computing Devices Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Chicago, Ill., October 9–14, 1960. Manuscript submitted June 7, 1960; made available for printing August 9, 1960.

K. K. Graupe is with the Boeing Airplane Company, Seattle, Wash.

determinant method, could serve as an independent check on other methods.

For the special case, a linear system of equations with constant coefficients, the matrix of which is positive definite, there is a very simple method of solving the problem.[3] But such conditions are seldom fulfilled.

Proofs are given which assume that ideal operational amplifiers are available for the convergence of the new methods. The success of the computing procedures is illustrated by several examples. Many applications of these computing procedures are described by Clymer.[1]

The Different Computing Methods

The computing methods will be described in several steps progressing from simple to more complex problems.

PROBLEM 1

One algebraic or transcendental equation for k unknown constant parameters x_ν is given ($\nu=1, 2, \dots k$)

$$f(t, x_1, \dots x_k)=0 \tag{1}$$

The computer is to solve for all parameters x_ν. If equation 1 were valid only for a certain instant of time t, then the solution would have k-1 degrees of freedom. But equation 1 is valid for a period of time and the computer exploits this fact; it takes a certain time to obtain the solution.

Now, f may depend on t by way of several given functions $g_1(t)$, $g_2(t)$, ... $g_m(t)$ and some may be derivatives of other $g_\mu(t)$. What kind of functions these are is not important immediately. Therefore the dependence of f on all the $g_\mu(t)$ is summarized in equation 1 by $f(t, \dots)$.

One of the parameters x_ν may be a function of time. Then the problem is defined as above and the circuit will work, if the speed of the computer is as high, as will be expressed later by condition 4.

SOLUTION OF PROBLEM 1

The computing system of Fig. 1 will produce a set of functions $X_\nu(t)$ which converge to one solution x_ν of equation 1. The functions $P_\nu(t)$ appearing in Fig. 1 are arbitrary, but positive, functions of time. In special problems one will dispose of the functions $P_\nu(t)$ so as to simplify the computer as much as possible. G_ν is the gain of the pertinent integrator ($G_\nu>0$). The symbol ϵ, which will be referred to as the error of the computer, is defined by

$$\epsilon=f(t, X_1, X_2, X_3, \dots X_k)$$

The following proof that $X_\nu(t)$ converges to one solution x_ν of equation 1 will make clear the operating principle of the circuit.

Differentiating the function $f(t, X_1, X_2, \dots X_k)$, one obtains

$$\frac{d\epsilon}{dt}=\frac{\partial\epsilon}{\partial t}+\sum_{\nu=1}^{k}\frac{\partial\epsilon}{\partial X_\nu}\dot{X}_\nu$$

and from Fig. 1

$$\dot{X}_\nu=-G_\nu P_\nu\left(\operatorname{sign}\frac{\partial\epsilon}{\partial X_\nu}\right)\epsilon \tag{2}$$

(The symbol sign y means $y/|y|$, i.e., $+1$ or -1 depending on the sign of y.) Hence

$$\frac{d\epsilon}{dt}=\frac{\partial\epsilon}{\partial t}-\epsilon\sum_{\nu=1}^{k}G_\nu P_\nu\left|\frac{\partial\epsilon}{\partial X_\nu}\right| \tag{3}$$

If one requires all functions $P_\nu(t)$ to be so large that

$$\left|\frac{\partial\epsilon}{\epsilon\partial t}\right|<\sum_{\nu=1}^{k}G_\nu P_\nu\left|\frac{\partial\epsilon}{\partial X_\nu}\right| \text{ when } |\epsilon|>|\epsilon_{\min}| \tag{4}$$

then

$$-\frac{d\epsilon}{\epsilon dt}>0 \quad \text{for} \quad |\epsilon|>|\epsilon_{\min}| \tag{5}$$

Obviously condition 4 cannot hold for $|\epsilon|<|\epsilon_{\min}|$, except for special functions $f(t, \dots)$, because of the denominator ϵ in the left side of the condition.

The inequality 5 expresses that ϵ will monotonically go to the "final" error $\epsilon_{\min}$. Afterward, ϵ will fluctuate between about $\pm|\epsilon_{\min}|$. Only in this practical sense is the output X_ν of the computer "converging" to the solution x_ν of equation 1.

Condition 4 requires that $\partial\epsilon/\partial t$ be bound. The computer will fail near poles of $\partial\epsilon/\partial t$. The other occasion where condition 4 cannot be fulfilled takes place around such instants t' where all $\partial\epsilon/\partial X_\nu$ vanish simultaneously. These instants are unique; otherwise equation 1 could not be solved for x_ν. Since X_ν is put out by an integrator, $X_\nu(t)$ will be a continuous function.

The term $\partial\epsilon/\epsilon\partial t$ is of the order of the main frequency in the $g_\mu(t)$'s. If the right side of condition 4 were a constant, its reciprocal could be called the time constant of the computer. Therefore condition 4 can be interpreted to mean that the time constant of the computer should be shorter than the main period in the given functions.

If, for instance, x_1 is an unknown function of time and $x_2, x_3, \dots x_k$ are unknown constants, it is advantageous to choose $G_2\approx G_3\approx \dots \approx G_k<<G_1$ in such ratio that all multiplier outputs in Fig. 1 have about the same signal level.

To use all multipliers to their full accuracy and for the best suppression of noise and disturbances, integrator gains should be made as low as possible provid-

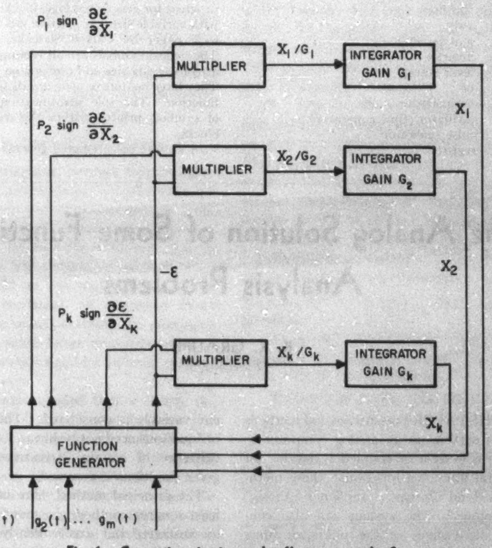


Fig. 1. Computing circuit to solve $f(t, x_1, \dots, x_k)=0$

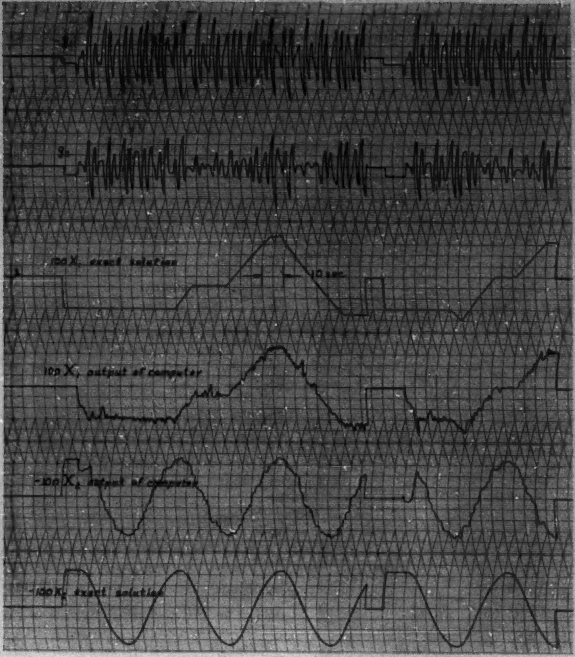


Fig. 2. Solution of $x_1\dot{g}_1+x_2g_1+g_2=0$, $G_1=G_2=1/\text{sec}$, $P_1=P_2=1{,}000$

ing that the multipliers do not overload. To increase the computer speed, and thereby the accuracy of the result, one should increase all P_ν, as long as the computer remains stable. According to condition 4 this procedure will decrease the final error $|e_{\min}|$.

In problem 1 the unknown quantities x_ν were assumed to be constant. In practical problems several parameters may vary slowly with time; then equation 1 does not determine the parameters sharply. Such a problem causes an additional loss of accuracy which cannot be avoided even with the fastest computer using the most accurate components. The following example represents such an ill-determined problem.

AN EXAMPLE OF PROBLEM 1

Fig. 2 is a record of the functions X_1 and X_2 which a computing circuit has put out as solutions of the equation

$$x_1\dot{g}_1+x_2g_1+g_2=0$$

The comparison of X_ν with the exact solutions x_ν demonstrates the accuracy obtained. Two computer runs are reproduced in Fig. 2. From the start of both runs one can read how long it takes for the computer to overcome wrong initial conditions: approximately 6 sec (seconds) or 2 cycles of the average frequency contained in the given functions g_1 and g_2.

This long response time of the computer is characteristic of the kind of problem. If the problem were well defined as, for instance, the one given in Fig. 3, one would obtain a much shorter response time.

PROBLEM 2

A more complex problem is to solve k coupled equations

$$f_\mu(t, x_1, x_2, \ldots x_k)=0 \qquad (6)$$

($\mu=1, 2, \ldots k$) for k unknown parameters or functions of time.

In general, two basically different methods are available to solve problem 2: the first will be called the method of least squares; the second, the determinant method. The first method is mentioned in an abstract by Brown,[2] but it was evidently applied only to linear systems. No description of the method was found. But because for more complex problems the least-squares method needs fewer components than the determinant method, the first method is proposed to solve this problem and the second as an independent check or for use in special problems.

LEAST-SQUARES-METHOD SOLUTION OF PROBLEM 2

It is convenient to abbreviate $e_\mu=f_\mu(t, X_1, X_2, \ldots X_k)$ and $e_{\mu\nu}=\partial e_\mu/\partial X_\nu$. The convergence of the computer output $X_\nu(t)$ to the exact solution $x_\nu(t)$ is obtained by setting

$$\dot{X}_\nu=-G\sum_{\mu=1}^{k} Q_\mu e_\mu e_{\mu\nu} \, (Q_\mu=\text{constant}>0) \qquad (7)$$

G is an integrator gain, chosen as small as possible for best filtering. Q_μ are high gains chosen to produce high speed and accuracy in the computer.

The right side of equation 7 may be considered as the νth component of the negative gradient of

$$\sum_{\mu=1}^{k} Q_\mu e_\mu^2$$

in the space of X_ν. Thus the point $(X_1, \ldots, X_k)$ is following a curve of steepest descent on the surface

$$\sum_{\mu=1}^{k} Q_\mu e_\mu^2$$

until it reaches a minimum.

The least-squares method can also solve problem 1. Then the computing method given for problem 1 includes the least-squares method, since equation 7 results from equation 2 when $P_\nu=Q_1|\partial e/\partial X_\nu|$ and when all G_ν are made equal.

Fig. 3 shows the solution of

$$x_1x_2-g_1(t)=0$$

$$2x_1+3x_2-g_2(t)=0$$

by the least-squares method. From the given functions $g_1(t)$ and $g_2(t)$, the computing circuit finds a solution X_1 and X_2. When the sign of $(3x_2-2x_1)$ changes, the computer can branch off to the other possible solution. In the example shown the computer just does that.

By applying a step change to x_1 or x_2 the response time of the computer was found to be about 0.1 second.

LEAST-MAGNITUDE-METHOD SOLUTION OF PROBLEM 2

The least-magnitude method is closely related to the least-squares method. By setting

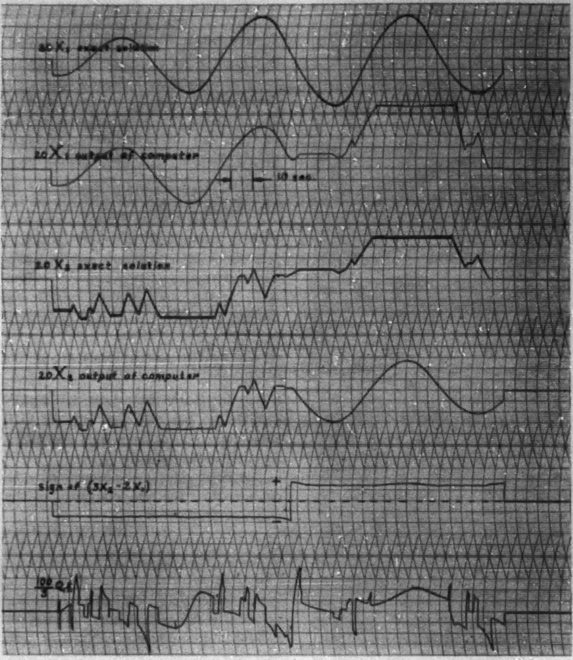


Fig. 3. Solution of $x_1x_2-g_1(t)=0$ $2x_1+3x_2-g_2(t)=0$, $G=1/\text{sec}$, $Q_1=200$, $Q_2=100$

$$\dot{X}_\nu=-G\sum_{\lambda=1}^{k} Q_\lambda\epsilon_{\lambda\nu}\text{ sign }\epsilon_\lambda \tag{8}$$

the point $(X_1, \ldots, X_k)$ will follow a curve of steepest descent on the surface

$$\sum_{\lambda=1}^{k} Q_\lambda|\epsilon_\lambda|$$

until it reaches a minimum. This is simpler than the least-squares method. Instead of k^2 multipliers which are necessary to compute $\dot{X}_\nu$ according to equation 7, only k^2 switches are needed to compute $\dot{X}_\nu$ according to equation 8.

The computing method given for problem 1 includes also the least-magnitude method when both methods are applied to problem 1.

When the set of equations 6 consists of m equations with k unknown parameters and when $m>K$, the problem is overdetermined. By extending the sums in equations 7 and 8 over μ from 1 to m, and by proportioning the Q_μ "justly," all available data can adequately contribute to the solution. Such simple participation of redundant data is not possible in the determinant method which is described in the following.

DETERMINANT-METHOD SOLUTION OF PROBLEM 2

Set

$$\dot{X}_\nu=-G\text{ sign }D\sum_{\lambda=1}^{k} P_\lambda(t)\epsilon_\lambda D_{\lambda\nu} \tag{9}$$

where $P_\lambda(t)$ are arbitrary positive functions of time, G is an integrator gain, chosen as low as possible for best filtering; $P_\lambda(t)$ are high gains chosen to obtain high speed and accuracy of the computer

$$D=\begin{Vmatrix} \epsilon_{11} & \ldots & \epsilon_{1k} \\ \cdot & & \cdot \\ \epsilon_{k1} & & \epsilon_{kk} \end{Vmatrix} \tag{10}$$

and $D_{\mu\nu}$ is defined by

$$D=\sum_{\nu=1}^{k} \epsilon_{\mu\nu}D_{\mu\nu}\text{ (for each }\mu=1, \ldots, k) \tag{11}$$

After integration $\dot{X}_\nu$ is fed back to the function generator. Instead of the sum

$$\sum_{\mu=1}^{k} Q_\mu\epsilon_\mu^2$$

each individual ϵ_μ^2 is minimized as shown in the following proof of convergence:

$$\frac{d\epsilon_\mu}{dt}=\frac{\partial\epsilon_\mu}{\partial t}+\sum_{\nu=1}^{k} \epsilon_{\mu\nu}\dot{X}_\nu$$

$$=\frac{\partial\epsilon_\mu}{\partial t}-G\text{ sign }D\sum_{\lambda=1}^{k}\sum_{\nu=1}^{k} \epsilon_{\mu\nu}D_{\lambda\nu}P_\lambda\epsilon_\lambda$$

According to a theorem on the development of determinants

$$\sum_{\nu=1}^{k} \epsilon_{\mu\nu}D_{\lambda\nu}=\frac{D}{0}\text{ when }\frac{\mu=\lambda}{\mu\neq\lambda}$$

which means that

$$\frac{d\epsilon_\mu}{dt}=\frac{\partial\epsilon_\mu}{\partial t}-GP_\mu|D|\epsilon_\mu \tag{12}$$

If one requires each $P_\mu(t)$ to be so large that

$$\left|\frac{1}{\epsilon_\mu}\frac{\partial\epsilon_\mu}{\partial t}\right|<GP_\mu|D|$$

when

$$|\epsilon_\mu|>|\epsilon_{\min}| \tag{13}$$

then

$$-\frac{1}{\epsilon_\mu}\frac{d\epsilon_\mu}{dt}>0$$

for

$$|\epsilon|>|\epsilon_{\min}|$$

and also

$$-2\epsilon_\mu\frac{d\epsilon_\mu}{dt}=-\frac{d}{dt}(\epsilon_\mu)^2>0 \tag{14}$$

But ϵ_μ^2 can decrease only until it reaches zero. So the determinant method minimizes each ϵ_μ^2 individually. It appears that minimizing each ϵ_μ^2 individually is a safer and more direct way to obtain a solution of equation 6 than minimizing only the one expression

$$\sum_{\mu=1}^{k} Q_\mu\epsilon_\mu^2$$

To save multipliers it is advantageous to set

$$P_\lambda\epsilon_\lambda=\text{constant} \tag{15}$$

where the constant has the same sign as ϵ_λ.

PROBLEM 3

Two equations

$$f_1(t, x_1, x_2, x_3)=0$$
$$f_2(t, x_1, x_2, x_3)=0 \tag{16}$$

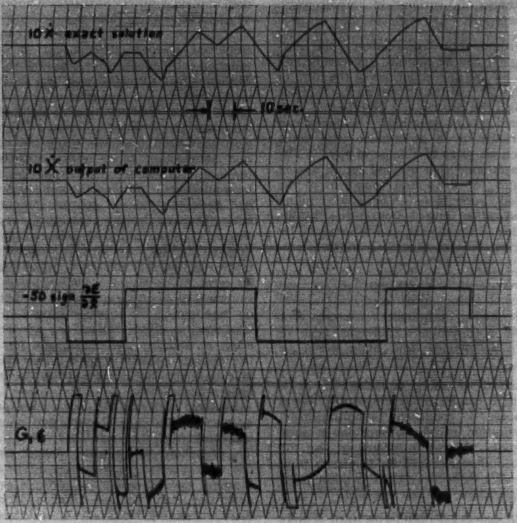


Fig. 4. Solution of $100\,g(t)-0.1x\dot{x}\sin 0.07t+10\dot{x}\cos 0.07t=0$, $G=1/6$ sec, $P=200$

are to be solved for three unknown constants or "slowly" varying parameters x_v. The determinant method of solving this problem rests on equalizing

$$\dot{X}_v=GE_vS \tag{17}$$

where

$$E_1=A_2(\epsilon_1\epsilon_{23}-\epsilon_2\epsilon_{13})+A_3(\epsilon_2\epsilon_{12}-\epsilon_1\epsilon_{22})$$

$$E_2=A_3(\epsilon_1\epsilon_{21}-\epsilon_2\epsilon_{11})+A_1(\epsilon_2\epsilon_{13}-\epsilon_1\epsilon_{23})$$

$$E_3=A_1(\epsilon_1\epsilon_{22}-\epsilon_2\epsilon_{12})+A_2(\epsilon_2\epsilon_{11}-\epsilon_1\epsilon_{21}) \tag{18}$$

$$S=A_1(\epsilon_{12}\epsilon_{23}-\epsilon_{13}\epsilon_{22})+A_2(\epsilon_{13}\epsilon_{21}-\epsilon_{11}\epsilon_{23})+A_3(\epsilon_{11}\epsilon_{22}-\epsilon_{12}\epsilon_{21}) \tag{19}$$

A_v ($v=1, 2, 3$) are three arbitrary functions of time, G is the integrator gain. In contrast the least-magnitude method would require

$$-\dot{X}_v=GQ_1\epsilon_{1v}\text{ sign }\epsilon_1+GQ_2\epsilon_{2v}\text{ sign }\epsilon_2 \tag{8A}$$

Thus for problem 3 the least-squares and the least-magnitude methods are much simpler and generally should be preferred.

The presentation given here is designed for analog computers but is applicable also to digital computers. The latter use will become mandatory when higher accuracy is required.

The procedures described do not change when given and unknown functions depend on more than one independent variable. In this case one can solve the equations by letting one variable run and keeping all others constant.

If the unknown functions appear together with their derivatives, the same methods are applicable. One must then find at first only the highest derivative of each unknown function.

The application of the methods described to differential equations eliminates the need to isolate the highest derivative, which may not always be possible in closed form. Two examples for this kind of problem are solved as shown in in Figs. 4 and 5.

Conclusions

1. Among the four demonstrated computation methods the least-magnitude method should be preferred in general because of its relative simplicity. If only one equation is given, the method given for problem 1 allows more freedom of design, and is therefore simpler and preferable.
2. Nonlinear problems can be solved. However, it may be necessary to feed additional information into the computer which selects, of several possible solutions, the one desired.
3. Two computation methods, the least-squares and the least-magnitude methods, can digest and weigh redundant data, thereby improving the accuracy of the computer. Without using especially quick computer components, a response time of 0.1 sec was obtained. Thus the computing methods shown in this report should be able to find "momentary" quantities continuously while a process is running. It is estimated that the use of redundant data also will decrease the response time of the computer.
4. These methods can solve ordinary differential equations, in which case it is not necessary to isolate the highest derivative.

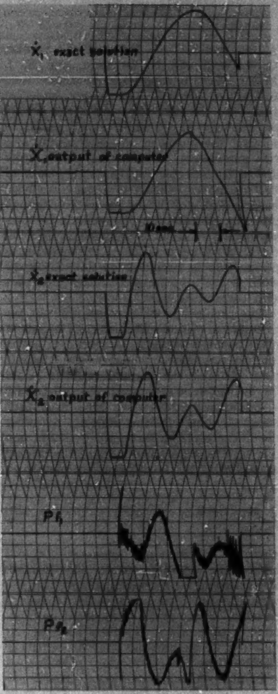


Fig. 5. Solution of $0=g_1(t)+\dot{x}_1\cos 0.06t-0.001x_1\dot{x}_2\sin 0.06t+0.1x_1$, $0=g_2(t)+{}_2\cos 0.06t+0.001x_1\dot{x}_1\sin 0.06t-0.1x_2$ by the determinant method. $G=0.2$/sec, $P_1=P_2=600$

References

1. Direct System Synthesis by Means of Computers, A. Ben Clymer. *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 77, 1958 (Jan. 1959 section), pp. 798–806.

2. The Stability of Feedback Solutions of Simultaneous Linear Equations, G. W. Brown. *Bulletin*, American Mathematical Society, Providence, R. I., vol. 53, pt. I, Jan. 1947, p. 61.

3. Linear Algebraic Systems and the REAC, Gebhardt, Landis. "Mathematical Tables and Other Aids to Computers," National Research Council, Wash., D. C., 1952, sec. 6, pp. 190–203.

Discussion

A. B. Clymer, T. F. Potts, W. C. Franke, C. R. Walli, G. N. Ornstein, and **B. J. Miller** (North American Aviation, Inc., Columbus, Ohio): The paper deserves study by all concerned with the design or analysis of dynamic experiments, in solving the problem of determining system parameters. Publication of this paper makes it desirable to report the work of the discussers which has been carried on independently of that of the author, though pursuing parallel developments. In presenting its relation to the general problem, together with extensions of the methods, a perspective is provided on Mr. Graupe's work and a step forward in an important area recognized.

One problem form is determining the parameters of systems characterized by one or more ordinary differential equations, given forcing function and dependent variable time histories. A general problem could be couched in mathematical terms, but only a simple case is discussed here, extensions being obvious. In the author's notation, the linear problem discussed is determination of n of k parameters x_ν of the system characterized by the differential equation:

$$\sum_{\nu=1}^{k} x_\nu g_\nu(t)=0 \tag{20}$$

where $n \leq k$, given all functions $g_\nu(t)$ and $k-n$ of the k parameters x_ν. One or more of $g_\nu(t)$ could be forcing functions and the remainder time derivatives of the dependent variable. Some or all of x_ν could be functions of time or g_ν. There is a duality in equation 20 permitting $x_\nu(t)$ to be either unknown coefficients, forcing functions, or some of both, and $g_\nu(t)$ to be accordingly known functions or constants. This class of problems denoted by equation 20 includes the general case of being given incomplete information concerning forcing functions, some information describing the system, and complete system response information, and one is to determine the missing information. An example would be determination of gust and buffet information and some aerodynamic parameters from flight test data.

At any time t, equation 20 is a condition on n unknowns. If the functions $g_\nu(t)$ are given throughout a time interval or at more than n instants, the n unknowns, if constants, are overdetermined. It is the large amount of redundant given information often available which characterizes the problem mathematically and underlies the existence of many distinct approaches.

A convenient procedure is to distinguish open-loop from closed-loop methods. Open-loop methods involve direct calculation of n coefficients from at least n independent conditions of equation 20. Closed-loop involves continuous or intermittent iterative procedures employing some error quantity, as in a servo. Open-loop methods are explicit; closed-loop, implicit.

To appreciate the richness of the field of open-loop methods, each given function can be visualized as $g_\nu(t, \omega)$, a surface plotted against time t, and frequency ω. The input data to the n conditions can be visualized as points or loci obtained by sampling or scanning the surfaces in various ways, such as:

1. Without regarding frequency, the functions $g_\nu(t)$ can be sampled repeatedly and stored for use in n equations of the form of equation 20. If the unknowns $x_\nu(t)$ are not constant, the process can be repeated with staggered sets of equations. Sampling can be periodic or in accordance with some criterion on rates of change of $g_\nu(t)$. The discussers tried this method on an analog computer for $n=3$ and found it to be reasonably successful. A periodic sampling interval Δt is limited from below by ill-condition of the matrix, and from above by resolution considerations for variable coefficients.

2. Short-term averages can be used as samples to reduce error from noise of frequencies higher than those of interest. For variable coefficients, a different type of error is introduced, and little would be gained.

3. Instead of using equation 20 at n instants, it can be differentiated $n-1$ times to obtain n conditions simultaneously. Differentiation introduces error due to noise for rapidly varying functions, or else the matrix is ill-conditioned for slowly varying functions. Either way there is a low practical upper limit on the value of n.

4. The curves $g_\nu(t)$ can be sampled at more than n times, and classical methods of regression analysis, multiple correlation analysis, least squares, etc., applied to obtain n conditions incorporating all of the sampled data. For variable coefficients x_ν, dynamic least squares[1] can be used.

5. It might be possible to sample the functions $g_\nu(t, \omega)$ along the frequency dimension to obtain n or more simultaneous conditions.

Closed-loop methods may be classified according to the error quantity employed. Implicit synthesis[1] uses an error ϵ_1 in the synthesized dependent variable, as do Meissinger,[3] and Margolis and Leondes.[4] Mr. Graupe and the discussers have shown the use of an error ϵ_2 in satisfying equation 20 by tentative values X_ν for the unknowns. Each method utilizes an error ϵ_i in some scheme such as "steepest descent"[5] down an even function of ϵ_i to drive the unknowns toward correct values.

Methods using error ϵ_1 in the dependent variable perforce use also a mathematical system model to synthesize continuously a tentative dependent variable. Margolis and Leondes call this model a "learning model," since in a sense it learns the proper values for coefficients. No explicit learning model is in the methods using error ϵ_2, though the whole circuit could be learning coefficients. The learning concept serves as a bridge to current activity in human learning simulation with digital computer programs, automata, etc.[6]

For one unknown coefficient, methods using dependent variable error ϵ_1 degenerate to implicit synthesis. Interestingly, both the author's and the discussers' investigations used implicit synthesis initially, but led finally to use of error ϵ_2 rather than ϵ_1. Independently of Clymer,[1] others in unpublished work have used implicit synthesis to generate a forcing function. Coincidentally the author and others have used implicit synthesis as their first approach when faced with nuclear reactor problems.

Meissinger, and Margolis and Leondes, employed steepest descent on an even function of dependent variable error ϵ_1. Mathematical formulation of the gradient used involves the concept of parameter sensitivity or perturbation"[7,8] as used in optimization studies.[3] Mr. Graupe's steepest descent methods using the equation satisfaction error ϵ_2 are apparently the first with significant generality to have been published. Earlier, the same principle had been used in fire control computers, analog sine and cosine generating circuits, and other specialized and isolated applications to continuously apportion corrections to two or more quantities contributing to an error. The author's methods may be summarized as continuous steepest descent on functions of the equation error ϵ_2. His least-squares method uses the square of error ϵ_2, and his least-magnitude method the absolute value of error ϵ_2. It is creditable that the chosen demonstration problems are nonlinear.

In these methods the function of error ϵ_2, when plotted against unknowns x_ν, may be visualized as a trough-like hypersurface. The trough rotates in some manner about the answer point as functions $g_\nu(t)$ vary. Steepest descent sends the process point down the side of the instantaneous trough. With adequate gains G_ν, the error decreases to $\epsilon_{\min}$. The process point's velocity component toward the answer point is small; eventually the answer point vicinity is reached. Independently, the discussers observed a 5-second convergence time constant for four unknowns in a linear third-order system.

Convergence rate of steepest descent can be increased by using sampled and held troughs. Various sampling criteria can be used, but periodic sampling is no better than continuous methods. It is conceivable that better performance could be obtained by accumulating products ϵg_ν algebraically for a time Δt to obtain a net correction vector directed at the answer point instead of along the gradient, repeating the cycle periodically or according to some error criterion. Running weighted averages of products ϵg_ν would permit a continuous solution.

Related methods can be derived by considering the equation error ϵ_2 as a sum of terms proportional to functions $g_\nu(t)$, each having a characteristic "signature." Appearance of the signature of a particular $g_\nu(t)$ in the error indicates that the tentative corresponding X_ν is too high. The coefficient could be decreased at a rate proportional to the amount of $g_\nu(t)$ in the error, estimated as some function of ϵg_ν. Other methods could employ both errors (ϵ_1 and ϵ_2) simultaneously or alternately to obtain the best features of both approaches. A blend of Meissinger and Graupe would be steepest descent on the sum of squares (or absolute values) of ϵ_1 and ϵ_2. New methods or variants could be conceived by substituting the concepts of linear programing,[3] steep descent,[5] or dynamic programing,[5] e.g., for steepest descent.

Considering again the class of problems described in conjunction with equation 20, the art of solution seems to deserve the name "analysis of dynamic experiments," to stand alongside the classical art of analysis of statistical experiments. The two arts, united by stochastic process theory or future mathematical developments in spectrostatistics, together with the art of experiment design, constitute a general field that could be termed "design and analysis of experiments." It would be desirable to extend the problem domain to include determination of partial differential equations, integral equations, etc., for systems having distributed parameters.

A sufficient number of distinct methods and formulations have been described to illustrate that the general problem of system parameter determination offers a wealth of research opportunities. Hopefully, readers will be stimulated to seek a unified mathematical description, help evaluate methods, and indicate those more flexible, convenient, rapid, and accurate.

Study areas to be addressed would be restrictions on forcing functions, optimization of forcing functions, methods development from stochastic process theory, restrictions on number and magnitude of linear system parameters to be determined, limitations on applications to nonlinear systems (e.g., time delays), limitations imposed by computer bandwidth, desirable specifications for computers and tape recorders, special problems associated with systems of equations, evaluation of best mathematical areas of application of each method, optimization of doctrine for dynamic experiments to take maximum advantage of method capability, and selection of highest priority areas of application. Mr. Graupe's opinions on some of these questions would be valuable guides to further research.

REFERENCES

1. See reference 1 of the paper.

2. LINEAR PROGRAMMING ON AN ELECTRONIC ANALOGUE COMPUTER, I. B. Pyne. *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 75, May 1956, pp. 139–43.

3. THE USE OF PARAMETER INFLUENCE COEFFICIENTS IN COMPUTER ANALYSIS OF DYNAMIC SYSTEMS, H. F. Meissinger. *Proceedings*, Western Joint Computer Conference, National Press, Palo Alto, Calif., 1960.

4. A PARAMETER TRACKING SERVO FOR ADAPTIVE CONTROL SYSTEMS, M. Margolis, C. T. Leondes. *Transactions*, Professional Group on Automatic Control, Institute of Radio Engineers, New York, N.Y., vol. PG AC-4, no. 2, Nov., 1959, pp. 100–11.

5. MODERN MATHEMATICS FOR THE ENGINEER (book), E. F. Beckenbach. McGraw-Hill Book Company, Inc., New York, N.Y., 1956, chaps. 11 and 18.

6. ACCOMPLISHMENTS IN HUMAN SIMULATION, A. B. Clymer. *Paper no. 60-AV-26*, American Society of Mechanical Engineers, New York, N.Y., June 1960.

7. ANALOG, DIGITAL, AND COMBINED ANALOG DIGITAL COMPUTERS FOR REAL-TIME SIMULATION, C. G. Blanyer, H. Mori. *Proceedings*, Eastern Joint Computer Conference, Institute of Radio Engineers, New York, N. Y., Dec. 1957, pp. 104–10.

8. A PERTURBATION TECHNIQUE FOR ANALOG COMPUTERS, L. Bush, P. Orlando. *Transactions*, Professional Group on Electronic Computers, Institute of Radio Engineers, vol. PGEC-8, no. 2, June 1959, pp. 218–21.

Kurt K. Graupe: I appreciate the comments of Mr. Clymer and his associates. Their discussion raises questions which I feel cannot be answered immediately. I am aware that the article itself leaves many questions open; e.g., the argument that the steepest descent down a surface must lead to a minimum on the surface is true only for stationary surfaces. On deforming and moving surfaces an additional condition must be fulfilled for the computer output to approach the minimum. This condition will be a mathematical formulation of the idea that the speed of the computer must be greater than the speed of deformation and translation of the error surfaces. Only the fact that examples have been tried which obtained "good" solutions (as good as actual computer components are) gave me the justification to publish this paper.

Quantitative Study of Low-Frequency Hysteresis Loops of Polarized Polycrystalline Barium Titanate

G. W. MARKS
NONMEMBER AIEE

D. A. HANNA
NONMEMBER AIEE

FERROELECTRICS belong to that group of dielectrics which are said to be nonlinear, but a nonlinear dielectric is not necessarily ferroelectric. The statement that a dielectric is nonlinear requires rather careful consideration.[1] If the temperature of a ferroelectric ceramic and also the external pressure on it are maintained constant, and if an electric field is applied which is below the saturation level, then the electric displacement D, or equivalently the polarization P, depends not only on the magnitude of the applied field but also on the past history and state of strain of the material. Linear dielectrics are characterized by the unique function $D(\omega)=\epsilon(\omega)E(\omega)$, which does not hold true for nonlinear dielectrics. This implies that when one is dealing with a nonlinear dielectric: (1) frequencies appear in $P(t)$, or in $D(t)$, which are not present in $E(t)$; and/or (2) if $E_1(t)$ produces $D_1(t)$ and E_2 produces $D_2(t)$, then $A_1E_1(t)+A_2E_2(t)$ does not produce $A_1D_1(t)+A_2D_2(t)$.

The nonlinearity of ferroelectrics is best exhibited, perhaps, by the P versus E hysteresis loops which characterize such materials. Much qualitative and descriptive information concerning these loops is to be found in the literature concerning ferroelectrics. For example, Heywang and co-workers studied the effect of grain size in the ceramic on the shape of hysteresis loops of unpolarized barium titanate ($BaTiO_3$).[2,3] They found that small grain size causes the domain configuration to be more markedly fixed and less responsive to field strength if time is taken into consideration. On long field-free storage of unpolarized polycrystalline samples, a domain configuration having a minimum of free energy results.

With modern facilities moderately accurate quantitative data can be deduced from oscillograms of hysteresis loops. It is the purpose of this paper to present, interpret, and discuss certain quantitative results obtained through the study of the hysteresis loops of polarized $BaTiO_3$ ceramics. Papers now in preparation will deal with the loops of unpolarized ceramics and with specific anomalies inherent in these loops.

Theory

It is of interest to study theoretically the maximum average directed polarization obtainable in a polycrystalline ferroelectric ceramic the domains of which have all possible orientations prior to polarization by an external field. The general case will be considered first.

Imagine a hemisphere of polarized ceramic of radius R and volume V as shown in Fig. 1. Assume that $\mathbf{P}$, the dipole moment per unit volume, is always

---

Paper 60-1285, recommended by the AIEE Basic Sciences Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Chicago, Ill., October 9–14, 1960. Manuscript submitted April 8, 1960; made available for printing September 20, 1960.

G. W. MARKS and D. A. HANNA are with the U. S. Navy Electronics Laboratory, San Diego, Calif.

The authors take the opportunity to thank Dr. R. N. Goss and his staff, of the Numerical Analysis Section, for carrying out the computations with the aid of electronic computers; Mrs. Marian Hicks who drafted the figures; Dr. R. D. Redin for helpful suggestions during the course of the work; and the Solar Manufacturing Company of Los Angeles, Calif., who kindly supplied the samples for investigation.

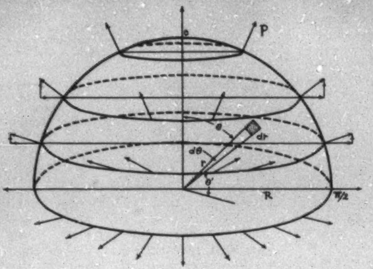


Fig. 1. A theoretical hemisphere of polarized ceramic $BaTiO_3$ of radius R and volume V

oriented in the radial direction, and that it has a constant magnitude and no components $\mathbf{P}_y$ directed opposite to the direction of the original polarizing field. We want to find the average $\bar{P}_y$ for this hemisphere.

Given, then, an element of volume dv in the hemisphere, if r is the radial distance of this element from the center of the hemisphere and θ is the angle this radius makes with a vertical plane

$$dv=(2\pi r\cos\theta)rdrd\theta \tag{1}$$

The vertical component P_y of the polarization vector is

$$\mathbf{P}_y=\mathbf{P}\cos\theta \tag{2}$$

Therefore the average value of $\bar{\mathbf{P}}_y$ for the hemisphere is given by

$$\bar{P}_y=\frac{\int_0^{\pi/2}\int_0^R P\cos\theta(2\pi r\sin\theta)rd\theta dr}{2/3\pi R^3}=P/2 \tag{3}$$

This result states that if the polarization vector is assumed to be constant in magnitude, then the average polarization normal to the basal plane of the hemisphere is equal to half this constant value.

Actually the magnitude of the vector P varies with direction and can be expected to have such a gaussian distribution that the maximum occurs in the direction of the original polarizing field. This distribution will depend on the degree of polarization, grain size, aging, etc. Suppose that P varies with direction in this manner:

$$P=P_Me^{-B\theta^2} \tag{4}$$

where P_M is the maximum dipole moment per unit volume and β is a constant. By way of illustration let $\beta=1$. Then equation 3 becomes

$$\bar{P}_y=\frac{3P_M}{R^3}\int_0^{\pi/2}\int_0^R e^{-\theta^2}\sin\theta \cos\theta d\theta r^2dr \tag{5}$$

$$=P_M\int_0^{\pi/2}e^{-\theta^2}\sin\theta\cos\theta d\theta \tag{6}$$

This integral was evaluated by plotting the integrand over the range from 0 to $\pi/2$ and determining the area. The result is $\bar{P}_y=0.26\,P_M$. The constant β determines the spread of the exponential curve, and so $\bar{P}_y$ can be properly evaluated only when β is estimated from experimental observations.

In tetragonal $BaTiO_3$ we must take into account the fact that any one of the three orthogonal crystallographic axes can become the polar axis. This means that the direction of polarization is limited to an angle of less than 90 degrees after the application of a d-c field of sufficient magnitude to orient the polar axes of all domains. The polar directions of the domains tend to take that crystallographic axis as polar axis which most nearly approximates the applied field direction.

The axial ratio for tetragonal $BaTiO_3$ is 1 to 1.01. Hence the angle between a diagonal of a unit cell and the polar axis is 54.5 degrees. This is the maximum angle that the polar axis of a domain in ceramic $BaTiO_3$ can make with the field direction providing that the polarizing field was of sufficient magnitude. We must now modify equation 3. It is seen that

$$\bar{P}_y=\frac{\int_0^{\theta^0}\int_0^R 2\pi Pr^2\sin\theta\cos\theta drd\theta}{\int_0^{\theta_0}\int_0^R 2\pi r^2\sin\theta d\theta} \tag{3A}$$

where $\theta_0=54.5$ degrees:

$$\bar{P}_y=0.79P$$

Thus the maximum possible value of the average field-directed spontaneous polarization is nearly 0.8 of that for a single domain crystal. However, it is always less than this because of incomplete orientation, relaxation, and aging after removal of the field.

Experimental Considerations

Disks of technical grade $BaTiO_3$ and $BaTiO_3$ with additives which had been manufactured under identical conditions for experimental work by a gas-firing process, machined to a diameter of one inch and a thickness of 100 mils, and provided with fired-on-silver electrodes, were polarized at approximately 45 volts per mil in hot silicone oil for about 30 minutes. Prior to use, these disks were aged at 80 C (degrees centigrade) for 24 hours and were then shelf-aged for a year. Compositions and average densities are given in Table I. Hereafter, identification of disks under investigation will be made by reference to (a), (b), (c), and (d) of this table. The calculated porosity of the (a) ceramics was 8.3%.

For the display of hysteresis loops the previously described circuit, utilizing capacitance voltage dividers, was employed at a frequency of 60 cps (cycles per second).[4] The output of this circuit was connected to the vertical and horizontal sweeps of a Hewlett Packard model *130A* oscilloscope. The rms values of these voltages were measured for calibration purposes with vacuum-tube voltmeters. Oscillograms were recorded by means of a Dumont oscillograph-record camera, type *297*, employing the Polaroid-Land photographic process. Enlargements of the oscillograms of about four diameters were made on glossy photographic paper. Transparent co-ordinate paper was then mounted over these enlargements and the necessary data read off. Finally, these data were converted to the desired units.

Samples for investigation were lightly clamped between two copper plates 1/4 inch thick, which served as high-voltage electrodes and aided in the rapid transmission of heat away from the sample. The unit was immersed in silicon transformer oil contained in a brass beaker. The latter was then placed in a constant temperature oil bath maintained at 25 C.

Polarized $BaTiO_3$ ceramics rise in temperature on the application of alternating fields of high strength because of the losses occurring and the low heat conductivity. In time, this temperature rise

Table I. Composition and Densities of $BaTiO_3$ Ceramics, Coupling Coefficients of Disks

Ceramic Designation	Weight, Per Cent			Coupling Coefficients, K_{15}	
	$PbTiO_3$ Added	$CaTiO_3$ Added	Density, gm/cm³	Before Experimentation	After Experimentation
(a)	0	0	5.56	0.31 0.31	0.30 0.31
(b)	2	0	5.54	0.28 0.29	0.28 0.29
(c)	4	0	5.54	0.28 0.28	0.28 0.29
(d)	12	8	5.22	0.20 0.20	0.19 0.19

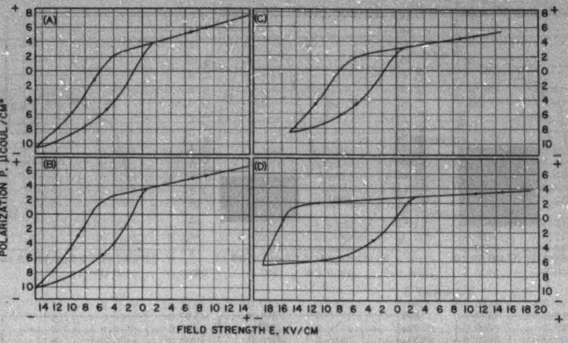


Fig. 2. Saturated hysteresis loops at 60 cps for polarized ceramic $BaTiO_3$ (polarization P in microcoulombs/cm² versus field strength E in kv/cm) for: A—$BaTiO_3$ (no additive). B—2% $PbTiO_3$ additive. C—4% $PbTiO_3$ additive. D—12% $PbTiO_3$+8% $CaTiO_3$ additive

results in depolarization of the sample. The copper electrodes used tend to minimize the effect. Further to ensure accurate results, readings were made and oscillograms were taken rapidly, usually in less than 15 seconds. By following this procedure only moderate warming of the samples was observed. Coupling coefficients of the samples were determined from admittance data before and after each experiment; representative results are contained in Table I. Essentially no change was observed. Thus, results were not altered by changes in the state of polarization during an experiment.

Meter readings of voltages applied to the horizontal and vertical plates of the oscilloscope were assumed to be rms values. This is not strictly correct, particularly for voltages referring to polarization or current as the distortion of the waveform is then greater. These distortions will be discussed later. The over-all accuracy of the results deduced from oscillograms was estimated to be ±10%.

Hysteresis Loops

Hysteresis loops at saturation showing the actual variation of polarization with field strength as deduced from the corresponding oscillograms for the four different compositions are presented in Fig. 2. The loop for ceramic (d) is shaped much like that for a single domain crystal of $BaTiO_3$. With increase in concentration of additive the remanent polarization P_r decreases somewhat and the coercive field strength E_c increases, as shown in Table II.

An anomalous loop or jog sometimes appears in the linear portion of a saturated P(or D) versus E loop for a polarized ceramic, i.e., when the direction of the field is in the direction of polarization but is declining in magnitude. Such a loop is observed in the oscillograms of Fig. 3(A). It first makes its appearance at a field strength of about 7 kv/cm (kv per centimeter). However, on the addition of lead titanate ($PbTiO_3$) this voltage becomes higher; for ceramic (d) it was above 16 kv/cm. With increase in maximum field strength applied to any given ceramic the position of this loop of the P versus E diagram moves out to higher voltages.

In an effort to find the source of this anomaly the operation of the display circuit was checked by replacing the ceramic with a high-voltage mica capacitor having a low dissipation factor. The oscillogram then observed was a straight line with positive slope when the vertical gain of the oscilloscope was zero. Voltage output to both the horizontal and vertical deflection plates of the oscilloscope was found to vary sinusoidally with time. Moderate changes in circuit constants other than the constant of the ceramic did not remove this loop. It must be concluded that this behavior is an inherent characteristic of the ceramic in conjunction with the described circuit.

As may be expected, when the barium titanate capacitor is in the circuit, the voltage output to the oscilloscope plates is not a fundamental sine wave. Examples are given in Figs. 3(A) through (D) for the described polarized ceramics. These figures show the variation of field strength and of polarization with fractional time as deduced from the corresponding oscillograms. The deviation from a pure sine wave is much greater for the polarization-versus-time curves. The anomalous loop occurring in the hysteresis loop now appears as the humps a and a'. The appearance of the hump in the field-versus-time curve as the field declines is equivalent to charging a condenser through a resistance when the declining field reaches a certain value, then adding the curve showing the variation of field strengths versus time on the condenser to the sine wave of the field applied to the ceramic. Perhaps a sudden switching of domains occurs in the ceramic with release of charge. The oscillograms for ceramics (a), (b), and (c) more nearly resemble each other than they do those for ceramic (d). For the latter the humps a and a' are less marked, and the polarization is about one-tenth that for the sample without additives.

It was of interest to make a Fourier analysis of these curves; this can be done readily by the use of electronic computers. Consider the equation

$$f_m(t/T)=\frac{A_0}{2}+\sum_{n=1}^{N} A_n \cos\left(\frac{2\pi mn}{M}\right)+\sum_{n=1}^{N} B_n \sin\left(\frac{2\pi mn}{M}\right) \qquad (7)$$

where

$$A_0=\frac{1}{M}\sum_{m=1}^{M} y_m$$

$$A_n=\frac{2}{M}\sum_{m=1}^{M} y_m \cos\left(\frac{2\pi mn}{M}\right)$$

and

$$B_n=\frac{2}{M}\sum_{m=1}^{M} y_m \sin\left(\frac{2\pi mn}{M}\right)$$

$A_0/2$ is the d-c component, and y_m refers to the magnitude of the polarization or of the field strength, as the case may be, at the fractional time t/T, where T is the period. For this work 100 ($=M$) equally spaced readings were taken from each respective curve in the range $t/T=0$ to 1. Thus $m=1, 2, 3, \ldots 100$ ($=M$) and $n=1, 2, 3, \ldots 50$ ($=N=M/2$). Partial

Table II. P_r and E_c as Determined From Saturated Hysteresis Loops for Certain $BaTiO_3$ Ceramics

Ceramic Designation	P_r, μcoul/Cm²	E_c, Kv/Cm
(a)	4.9	5.9
(b)	5.1	7.2
(c)	4.6	7.9
(d)	4.1	15.4

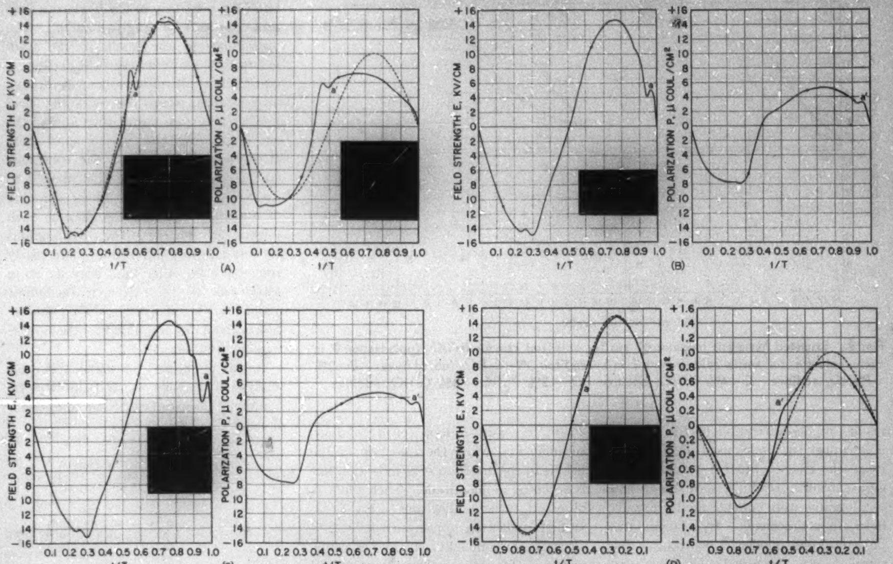


Fig. 3. Variation of field strength E and polarization P with fractional time t/T respectively for polarized $BaTiO_3$ ceramic disk. A—$BaTiO_3$. B through D—Additives as follows: B—2% $PbTiO_3$. C—4% $PbTiO_3$. D—12% $PbTiO_3$+8% $CaTiO_3$; pure sine waves (dashed curves) introduced for comparison

results of the computations are presented in Table III. A_n and B_n are the Fourier coefficients. If u is the column number and v is the row number in this table, then the coefficient number $n = u + (v-1)5$.

An examination of Table III shows that in all cases the d-c component was quite small. The variation of the field strength with time is largely represented by those frequencies having coefficients in the first row, whereas the variation of the polarization is primarily represented by frequencies having coefficients in the first and second rows.

Energy Loss

The area enclosed by a ferroelectric hysteresis loop represents energy loss per cycle. This loss depends upon the chemical composition and domain structure of the material, and on the temperature, pressure, and frequency of the applied field. In a polarized ferroelectric ceramic the loss occurs almost entirely when the direction of the applied field is opposite to the direction in which the ceramic was polarized.

In Fig. 4 are shown plots of energy loss in calories per gram per cycle versus maximum field strength in kv/cm for the polarized ceramics previously discussed. These losses were calculated from calibration data and planimeter measurements of the area of the loops on enlarged oscillograms. It is observed from the figure that at low maximum field strengths the losses are negligible and that there is a field strength E_0, above which the losses rise rapidly. Above E_0 it was found that the empirical expression

$$l = a(E_m - E_0)^2 \tag{8}$$

satisfactorily represents the experimental results. In this equation l is the loss in calories per gram per cycle, E_m is the maximum field strength and a and E_0 are constants. E_0 increases with increase in added percentage of $PbTiO_3$. Deviation from the empirical curve occurs when saturation sets in, for thereafter the polarization varies linearly with field strength as for an ordinary dielectric. Constants E_0 and a for different compositions are given in Table IV. Since energy storage in an ordinary linear dielectric varies as E^2, perhaps it is to be expected that the losses in a ferroelectric also would vary as E^2.

Differential Permittivity

The permittivity of $BaTiO_3$ varies considerably with electrical field strength, temperature, and mechanical stress. Among the properties with which we should be concerned in our discussion of hysteresis loops is the differential permittivity ϵ_d. This is given by the relation $\epsilon_d = 1 + 4\pi\eta_d$; $\eta_d = (\partial P/\partial E)$ is defined as the differential susceptibility at any point on the virgin curve or on the hysteresis loop for a ferroelectric single crystal, a polarized or an unpolarized ferroelectric ceramic.[6] The η_d is written in partial form because it is dependent on temperature, frequency, field strength, mechanical strain, chemical composition, and the previous history of the ferroelectric material. In this work the virgin curve was not recorded. Only the differential permittivity determined from the loop for polarized ceramics will be discussed. The η_d is high for $BaTiO_3$, so that within the limits of experimental error $\epsilon_d = 4\pi(\partial P/\partial E)$.

Studies were made of the variation of differential permittivity with field strength at increasing maximum field strengths for

Table III. Field Strength and Polarization Versus Fractional Time

Fig. No.	Row No.	Column No.										
		1		2		3		4		5		
		A_n	B_n	A_n	B_n	A_n	B_n	A_n	B_n	A_n	B_n	A_0
						E Versus t/T						
3(A)	1	0.4763	−14.6810	0.0767	0.3319	0.1791	0.0715	−0.2102	0.4425	−0.3574	−0.0898	−0.0900
	2	−0.2251	0.0470	0.0076	−0.3694	−0.0167	0.1277	0.3149	−0.0999	−0.2843	0.2123	
	3	0.2722	−0.0250	−0.3275	−0.1832	0.2753	0.2133	−0.0454	−0.2417	−0.0403	0.2468	
	4	0.0486	−0.1514	−0.0869	0.0790	0.1060	−0.0607	−0.0972	0.0147	0.0585	0.0125	
	5	−0.0378	−0.0160	0.0207	0.0048	−0.0168	−0.0096	0.0219	0.0061	−0.0240	−0.0060	
						P Versus t/T						
	1	−3.3715	−9.0258	2.2835	−1.2689	−0.2117	−0.3641	0.8101	−0.8722	0.4379	0.2099	−0.0775
	2	0.1458	−0.2866	0.3426	0.1980	−0.1076	0.0217	0.1765	0.0787	−0.1750	0.0792	
	3	0.0534	−0.0387	−0.0986	0.0478	0.0141	−0.0691	−0.0316	0.0366	−0.0089	−0.0539	
	4	0.0070	0.0301	−0.0268	−0.0198	0.0167	−0.0019	−0.0272	−0.0010	0.0153	−0.0208	
	5	−0.0144	0.0130	−0.0041	−0.0134	−0.0068	0.0047	−0.0043	−0.0109	−0.0110	−0.0020	
						E Versus t/T						
3(B)	1	−0.2512	14.7783	−0.0404	0.1413	0.0555	−0.0331	−0.1031	0.0507	0.1754	−0.0766	0.0385
	2	−0.0440	0.0049	0.0752	0.0172	−0.0012	−0.0337	0.0156	0.0588	0.0298	−0.0297	
	3	0.0129	−0.0230	0.0229	0.0144	0.0031	−0.0088	0.0023	0.0216	0.0137	−0.0043	
	4	−0.0244	0.0004	0.0133	0.0193	−0.0020	−0.0230	−0.0025	0.0078	0.0126	−0.0121	
	5	−0.0084	0.0019	0.0066	0.0107	−0.0001	−0.0075	−0.0046	0.0134	0.0050	−0.0040	
						P Versus t/T						
	1	0.0822	0.9425	−0.1371	−0.0773	0.0453	−0.0414	−0.0103	0.0022	0.0285	−0.0125	0.0024
	2	−0.0141	0.0239	0.0000	−0.0137	−0.0013	0.0078	0.0006	−0.0089	0.0050	0.0072	
	3	−0.0046	−0.0031	0.0038	0.0019	−0.0029	0.0000	0.0024	−0.0002	−0.0016	0.0004	
	4	0.0007	0.0000	0.0005	0.0001	−0.0003	−0.0004	−0.0012	0.0002	−0.0002	0.0005	
	5	0.0001	−0.0006	0.0008	0.0005	−0.0004	0.0003	0.0005	0.0000	−0.0007	−0.0009	
						E Versus t/T						
3(C)	1	0.3619	14.6600	−0.0832	0.2323	0.0090	−0.1044	−0.1366	0.2758	0.3477	−0.1009	0.0165
	2	−0.0361	0.0307	0.0703	0.1859	0.1892	0.0761	0.0230	0.0422	0.0124	0.2327	
	3	0.0936	0.1802	−0.0085	0.1599	−0.0960	0.1370	−0.1012	0.1712	−0.0975	0.0506	
	4	−0.1094	0.0252	−0.1075	−0.0072	−0.0753	−0.0099	−0.0524	−0.0252	−0.0336	−0.0346	
	5	−0.0386	−0.0163	−0.0212	−0.0146	−0.0097	−0.0183	−0.0208	−0.0307	−0.0160	0.0010	
						P Versus t/T						
	1	−1.7564	6.4286	1.3495	1.2926	0.5417	−0.1104	−0.0902	0.6803	0.4732	0.2052	0.0165
	2	−0.3528	−0.0103	−0.0090	0.2491	0.0887	0.1262	−0.0206	0.0310	−0.0557	0.0718	
	3	−0.0139	0.0748	−0.0043	0.0135	−0.0450	0.0069	−0.0197	0.0326	−0.0182	−0.0071	
	4	−0.0367	−0.0191	0.0000	−0.0116	0.0108	−0.0242	−0.0059	−0.0018	0.0149	0.0211	
	5	0.0086	−0.0019	−0.0173	−0.0023	0.0047	0.0014	0.0114	−0.0177	0.0040	−0.0030	
						E Versus t/T						
3(D)	1	0.7305	14.5776	−0.0331	0.3026	−0.0472	−0.0296	−0.1304	0.2011	0.3681	−0.0797	−0.0445
	2	0.0431	−0.0429	0.0699	0.1858	0.2699	0.1325	0.0533	0.0853	0.0311	0.2465	
	3	0.1805	0.2079	0.1161	0.1684	0.0227	0.2588	−0.0252	0.3315	−0.0716	0.2421	
	4	−0.1528	0.1829	−0.1882	0.1218	−0.1725	0.0869	−0.1289	0.0345	−0.1122	−0.0101	
	5	−0.0876	−0.0149	−0.0489	0.0154	−0.0360	0.0103	−0.0500	0.0190	−0.0620	0.0190	
						P Versus t/T						
	1	−1.9110	5.8872	1.3827	1.6488	0.2454	−0.0956	−0.1984	0.6334	0.4510	0.4245	0.0160
	2	0.1077	0.0050	−0.0637	0.1828	0.0975	0.1865	0.0477	0.0667	−0.0258	0.0869	
	3	−0.0163	0.0887	−0.0252	0.0566	−0.0352	0.0364	−0.0387	0.0137	−0.0439	0.0075	
	4	−0.0161	0.0011	−0.0034	−0.0157	−0.0151	−0.0118	0.0032	−0.0066	0.0067	−0.0125	
	5	−0.0056	−0.0121	−0.0042	−0.0066	0.0035	−0.0179	−0.0067	−0.0199	−0.0010	−0.0070	

Coefficients A_n and B_n in the Fourier series: $f(m)\left(\frac{t}{T}\right)=\frac{A_0}{2}+\sum_{n=1}^{N} A_n \cos\left(\frac{2\pi mn}{M}\right)+\sum_{n=1}^{N} B_n \sin\left(\frac{2\pi mn}{M}\right)$ **for *a*: ceramic $BaTiO_3$; *b*: ceramic $BaTiO_3$ containing 8% $CaTiO_3$+12% $PbTiO_3$ additive; *c*: ceramic $BaTiO_3$ containing 2% $PbTiO_3$ additive; *d*: ceramic $BaTiO_3$ containing 4% $PbTiO_3$.**

ceramics (a), (c), and (d). Differential permittivities were determined by measuring the angle of slope at points on a given hysteresis loop, then making the necessary calculations based on the calibration of the oscillogram; for results see Figs. 5 through 7. Inserts of the corresponding oscillograms are enclosed with each graph. In all cases the direction of rising positive field and the direction of polarization are the same in the first quadrant of the oscillogram. Differential permittivities are always positive in sign.

These curves are representative rather than absolute for, in general, no two ceramics which are identical in size and composition yield precisely the same oscillograms. The degree of polarization, represented by the coupling coefficient, is seldom exactly the same for two samples from the same lot. In some cases duplicate studies were made, but marked variations were not observed. However, the results given are not mean determinations but are, in each case, those for a single disk.

Some interesting deductions can be made from an examination of these curves which show the rate of change of polarization with field strength. Even at relatively low maximum field strengths the deviation from the linearity which characterizes ordinary low-loss dielectrics is more apparent than is indicated by the original oscillograms; see curve *A* in Figs. 5 through 7. Linearity, of course, often occurs when the direction of the field is in the direction of polarization. Although this portion of some of the oscillograms is not strictly linear, it has been so interpreted largely because of the inherent limitations in obtaining data from oscillograms; see the inserted oscillograms of Figs. 5 through 7.

All these curves are similar in shape. Choose any representative curve (e.g., *E*, Fig. 5) and consider the differential permittivity corresponding to maximum negative field strength. As the field declines negatively the differential permittivity rises to a peak *a* where $(\partial\epsilon_d/\partial E)=0$. It is found that with increase in the maximum applied field this peak rises in magnitude rather slowly as saturation sets in. An S-shaped curve is the result, as shown in Fig. 8. With increase in maximum applied field the position of this peak moves closer to the axis of ordinates; see Table V. After peak *a* has been reached, the differential permittivity declines rapidly and nearly linearly to the constant value ϵ_{d1} which describes

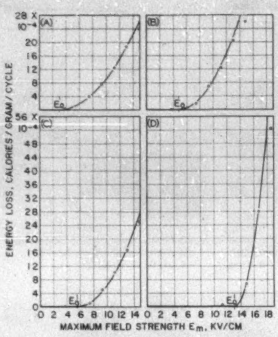


Fig. 4. Dependence of energy loss in calories/gram/cycle on maximum field strength in kv/cm for $BaTiO_3$ ceramics. Additives are A—None. B—2% $PbTiO_3$. C—4% $PbTiO_3$. D—12% $PbTiO_3$+8% $CaTiO_3$. Experimental data indicated by plotted points. Curves represent equation $l=a\ (E_m-E_0)^2$

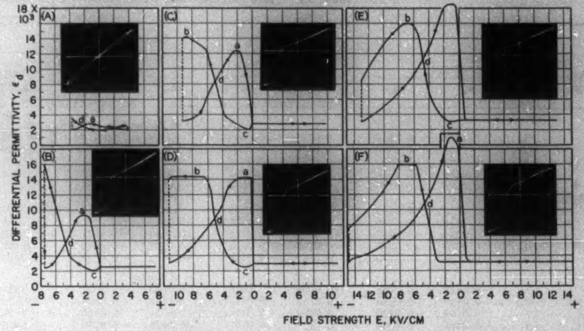


Fig. 5. Variation of differential permittivity ϵ_d with field strength E at increasing maximum field strengths for polarized $BaTiO_3$ ceramic. Refer to Table V for values of maximum field strengths. No additives

the polarized ceramic. For well-polarized samples ϵ_{d1} commences at or very near zero field and then remains essentially constant to maximum positive field strength. On decrease of positive field ϵ_{d1} remains about the same until reversal of field occurs, whereafter there is sometimes a decline to a small minimum at low negative field strength, e.g., see curve *D*, Fig. 5(A). The differential permittivity then rises with increase in the negative field, and at higher field strengths reaches a maximum value (peak *b*) before it commences to decline to the tip of the loop. It finally returns to the original value and the cycle is repeated. Also at peak *b* and minimum *c*, $(\partial\epsilon_d/\partial E)=0$. The characteristic differential permittivity ϵ_{d1} rises a little in magnitude with increase in maximum applied field; see Table V.

It is observed in all these graphs that there is a point where the portion of the curve for rising negative field strength intersects that for declining negative field strength. This corresponds to opposite points on the hysteresis loop where the slopes are the same. The differential permittivity at this point of intersection will be designed as ϵ_{d2}. With increase in maximum negative field ϵ_{d2} increases; see Table V.

Normal Permittivity

The normal or over-all permittivity ϵ_n is defined as $\epsilon_n=1+4\pi\eta$ where $\eta=P/E$. In this expression *P* is the polarization observed at any point on a hysteresis loop for a stated field strength *E*. Normal

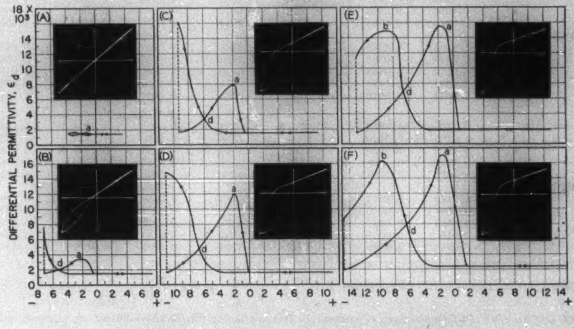


Fig. 6. See caption of Fig. 5. Additive: 4% $PbTiO_3$

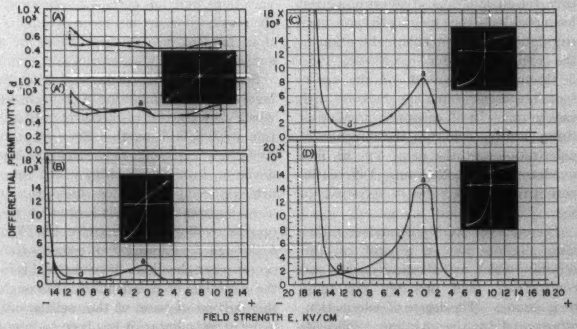


Fig. 7. See caption of Fig. 5. Additive: 12% $PbTiO_3$+8% $CaTiO_3$

Table IV. Constants' Loss in Calories

Polarized $BaTiO_3$ Ceramic Designation	Loss in Calories/Gram/Cycle at 25 C	
	Constant a (×10⁻⁴)	Constant E_0, Kv/Cm
(a)	0.186	3.0
(b)	0.320	4.8
(c)	0.309	5.5
(d)	2.10	13.0

Constants in empirical equation $1=(E_m-E_0)^2$.

Table V. Variation of Permittivities for Polarized $BaTiO_3$ Ceramics

Curve Designation	Maximum Field Strength, Kv/Cm	Characteristic Differential Permittivity ϵ_{d1} (×10³)	Field Strength at Intersection d, Kv/Cm	Differential Permittivity ϵ_{d3} at Intersection d (×10³)
		Ceramic (a), Fig. 5(A)		
A	3.75	0.0	−2.3	2.5
B	7.46	2.5	−4.7	5.7
C	9.28	2.8	−5.0	8.4
D	11.13	2.9	−5.2	8.8
E	13.0	3.3	−4.8	11.3
F	14.84	3.2	−4.7	10.8
		Ceramic (c), Fig. 5(B)		
A	3.71	1.4	−2.1	1.5
B	7.42	1.5	−5.1	2.0
C	9.28	1.5	−5.9	3.4
D	11.13	1.6	−6.8	4.9
E	13.0	1.8	−6.7	6.9
F	14.84	2.4	−6.4	7.8
		Ceramic (d), Fig. 5(C)		
A	11.2	0.5	−2.4	0.6
B	14.84	0.6	−10.2	0.8
C	16.70	0.6	−11.0	1.0
D	18.56	0.7	−12.4	1.6

permittivities for saturated loops were determined from data representing the curves of Fig. 2. Results for the four different compositions are plotted in Fig. 9.

Let us examine these curves. They are much the same for ceramics (a), (b), and (c). Except at field strengths in the range −1 to +1 kv/cm, ϵ_n ranges from about $0.9^{7}6$ to $1.0^{7}5$. The linear portion of the hysteresis loop where saturation occurs yields, of course, a hyperbolic variation of normal permittivity with field strength. At negative field strengths two branches are obtained: branch *a* corresponding to the upper portion of the hysteresis loop *a* (Fig. 9) and branch *b* corresponding to the lower portion of the hysteresis loop *b*. Both branches approach an ϵ_n of 0 at very low field strengths. Curves *b* have a maximum which lies between values of ϵ_n of $1.0^{7}10$ to $1.0^{7}13$ at −4 to −6 kv/cm.

Current Density Versus Field Strength Diagrams

If the capacitor in series with the ferroelectric sample in the circuit for the display of hysteresis loops is replaced by a resistor, then the figure observed on the oscillograph screen is one in which the abscissa is proportional to the applied field and the ordinate proportional to the current in the ferroelectric.[6] Then, one can study the variation of current density *J* with field strength *E*. For this work the 7.2-μf (microfarad) capacitor of the display circuit was replaced with a 970-ohm noninductive resistor.

The (*J*, *E*) diagram for an ordinary linear dielectric, when a sinusoidal potential is applied, is an ellipse. If, during the cycle, nonlinear polarization changes occur in a dielectric, then changes in permittivity result with consequent changes in impedance so that the diagram is no longer such a curve. When the average permittivity falls to a minimum the current rises to a maximum. The diagram for an unpolarized ceramic is centrosymmetrical about the origin, but that for a polarized sample is not.[3]

Oscillograms were taken at increasing maximum potentials up to saturation for the described ceramic disks. Current density *J* in amp/cm² (amperes/cm²) versus field strength *E* in kv/cm were deduced therefrom. Representative data for a $BaTiO_3$ disk containing no additives are plotted in Fig. 10. It is seen that at a low maximum field strength of 0.745 kv/cm the resulting (*J*, *E*) diagram approximates the ellipse

$$\frac{E^2}{(0.745)^2}+\frac{J^2}{(0.40\times10^{-4})^2}=1 \qquad (9)$$

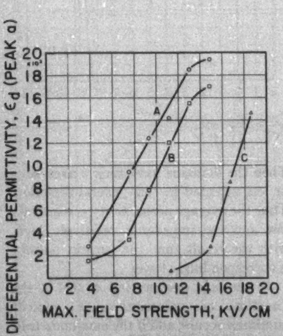


Fig. 8. Dependence of maximum differential permittivity (peak a) on field strength for $BaTiO_3$. Additives as follows: A—$BaTiO_3$, no additive. B—4% $PbTiO_3$. C—12% $PbTiO_3$ + 8% $CaTiO_3$

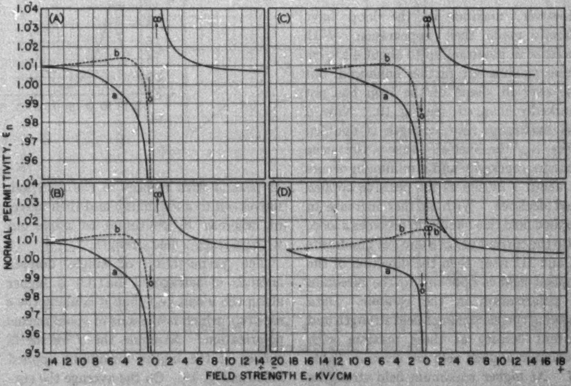
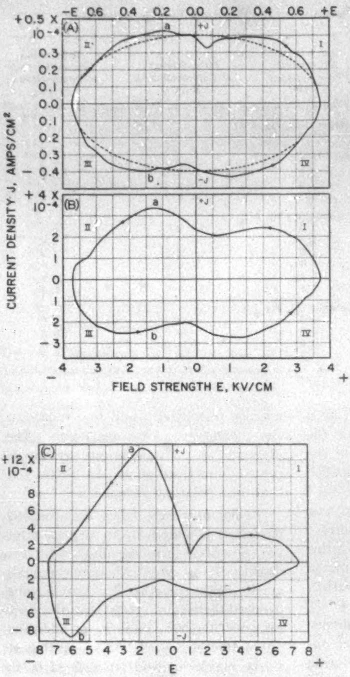


Fig. 9. Dependence of normal permittivity ϵ_n on field strength E in kv/cm. Additives as follows: A—$BaTiO_3$, no additive. B—2% $PbTiO_3$. C—4% $PbTiO_3$. D—12% $PbTiO_3$ + 8% $CaTiO_3$



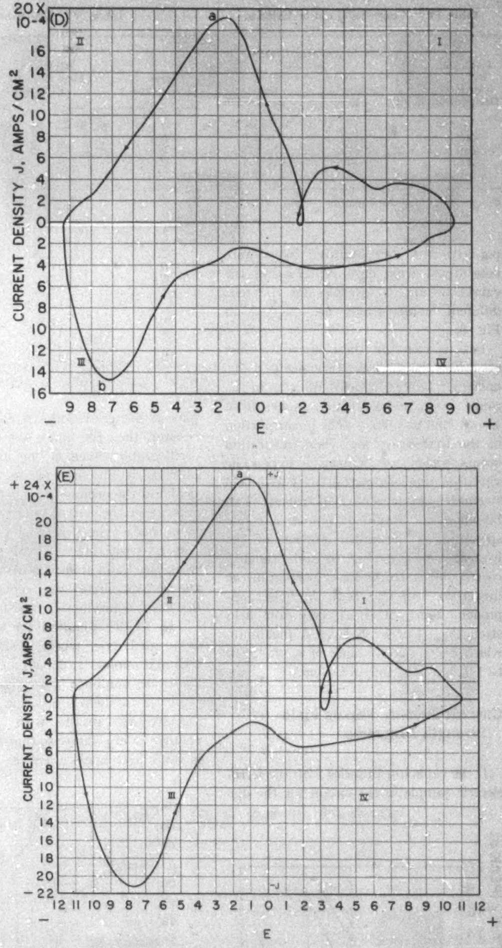


Fig. 10. Variation of current density J in amp/cm² with field strength E in kv/cm for increasing maximum field strength as follows: A—0.745. B—3.72. C—7.45. D—9.20. E—11.2

However, maxima in the loop have already made their appearance. With increase in potential up to about 4 kv/cm these maxima are further enhanced, but the over-all shape of the loop has not markedly changed; see Fig. 10. Now, if the maximum field strength is doubled (Fig. 10), the over-all shape of the curve changes markedly and a minimum occurs at about 1 kv/cm as the positive applied field declines. With further increase in applied field this minimum becomes a loop. Also the current peaks *a* and *b* in quadrants II and III rise; then the direction of the applied field is opposite to the direction of polarization in the ceramic. Data showing the variation of field strength and current density at these maxima with maximum applied field are given in Table VI. At higher maximum field strengths the current densities at these peaks, J_a and J_b respectively, change linearly with E_M. As seen by Table VI, the position of peak *a* with respect to the field axis moves out to a maximum and then moves in, whereas E_b continues to move out with increasing field strength.

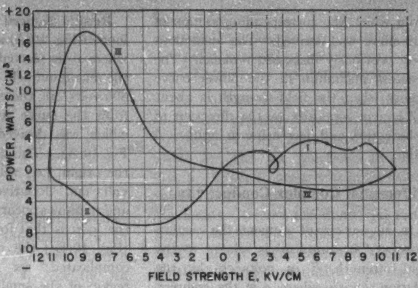
The permittivity of single domain crystals of tetragonal $BaTiO_3$ is higher along the *a* axis (ϵ_a) than it is along the *c* axis (ϵ_c), which is the axis of spontaneous polarization.[7] Examine the *J* versus *E* curves at the higher maximum applied fields in Fig. 10. On the average the current density is lower when the direction of the field is in the direction of polarization of the ceramic (quadrants I and IV). Then the domains have their *c* axes more or less oriented in this general direction. Thus the average permittivity is relatively low and the impedance relatively high, with the result that the current density is correspondingly lower.

As the field declines but is still in the direction of polarization, a current minimum occurs, and if the maximum field strength has been sufficiently high this minimum results in the formation of a small anomalous loop as previously pointed out. The loop corresponds to the regions *a* and *a′* of the field-versus-time

Table VI. Variation of Current Density and Field Strength

Maximum Applied Field E_M, Kv/Cm	Peaks, Fig. 10			
	a		b	
	E_a, Kv/Cm	J_a, Amp/Cm² (×10⁻⁴)	E_b, Kv/Cm	J_b, Amp/Cm² (×10⁻⁴)
0.745	−0.2	0.43	0.32	0.4
3.72	−1.30	3.36	−2.3	2.46
7.45	−1.75	13.2	−6.1	8.80
9.20	−1.60	19.1	−7.2	14.7
11.2	−1.10	24.5	−7.7	21.2

Fig. 11. Energy storage and return as derived from Fig. 10(E) for a polarized $BaTiO_3$ capacitor with no additives



and polarization-versus-time curves respectively of Figs. 3(A) through (C). The loop arises gradually with increase in field strength and first appears as a minimum in the curve. Only a point of inflection is observed in the field strength versus time curve. Although the applied field is declining, the resultant field of the ceramic falls and then rises again with time; see Fig. 3(A). In the course of the anomalous loop the current declines to zero then rises to a maximum in the opposite direction and again declines to zero, after which it continues to rise positively with declining field strength; see Fig. 10. The size and abscissas of this loop increase with increase in maximum field strength, which shows it to be time-dependent.

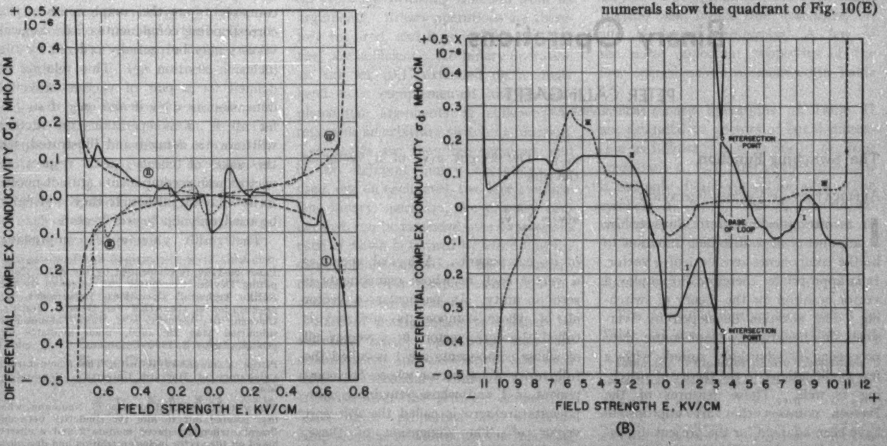
The explanation for the occurrence of this loop probably lies in the appearance of charge and its change with time and field strength at domain boundaries. At the contact surfaces of two arbitrarily oriented domains the limitations concerning the continuity of the normal components of electric displacement D_h are not met, so a surface charge must appear at the interface. Consider the relationship for the charge density between two media across which a current J_n is flowing:

$$\rho_s = J_n\left(\frac{\epsilon_1}{\sigma_1} - \frac{\epsilon_2}{\sigma_2}\right) \qquad (10)$$

where ϵ_1, ϵ_2 and σ_1, σ_2 are the permittivities and conductivities respectively of the two media.[8] The permittivity of ferroelectrics is not only anisotropic but also is dependent on field strength. If two domains are rather differently oriented with respect to the field, then ϵ_1 and ϵ_2 will differ with respect to this field and will change with it, however, often with a time lag and with growth and decline of the domains. The σ_1 and σ_2 would also vary somewhat with field strength, but this effect can be expected to be quite secondary to the permittivity change. This discussion refers, of course, to current through the media. The d-c component is small. However, the magnitude of the alternating current is dependent on the permittivities and accumulated charge.

When $\epsilon_1/\sigma_1 = \epsilon_2/\sigma_2$, the charge between the boundaries of the media drops to zero. When $\epsilon_1/\sigma_1 > \epsilon_2/\sigma_2$, then ρ_s/J_n is positive, whereas when $\epsilon_1/\sigma_1 < \epsilon_2/\sigma_2$, it is negative. Statistically the over-all charge density at the interfaces changes, with the result that there is a change in internal field and consequently in current density.

The rate of energy storage and return in watts per cubic cm versus field strength in kv/cm for a polarized $BaTiO_3$ capacitor (with no additives) under saturated conditions as derived from Fig. 10(E) is shown in Fig. 11. Roman numerals show the quadrant of Fig. 10(E)



(A) (B)

Fig. 12. Variation of differential complex permittivity in mhos/cm with field strength in kv/cm for $BaTiO_3$ disk. A—Dash-dot curve refers to ellipse of Fig. 10(A) from which it was derived; roman numerals refer to quadrant numbers of Fig. 10(A). B—Intersection point is that of loop in Fig. 10(E); roman numerals refer to quadrant numbers of Fig. 10(E)

from which the corresponding section of the curve was derived. The anomalous loop previously discussed remains. The greatest maximum in the power curve occurs when the field strength is about 8.7 kv/cm, opposite to the direction of polarization and decreasing in numerical value.

Differential Complex Conductivity

In an a-c circuit element the current density J in amp/cm^2 is related to the field strength E in volts/cm by the relation $J=\sigma E$, where σ is the complex conductivity in mhos/cm. Since a polarized $BaTiO_3$ ceramic condenser is a nonlinear circuit element, σ varies with E. The σ can be determined at any point on the J versus E loop. Because the dissipation factor of these condensers is quite small, the complex conductivity is largely imaginary.

The slope at any point on the J-versus-E loop is the rate of change of the current density with respect to the field at some particular value of the rising or declining field; thus $(dJ/dE)=\sigma_d$. The σ_d can be defined as the differential complex conductivity.

Results deduced from the low-voltage loop of Fig. 10(A) are shown in Fig. 12(A) where σ_d is plotted versus E. Although the maximum field strength was less than 800 volts/cm, there is a marked deviation from the curve for a linear condenser which was derived from the ellipse of Fig. 10(A). This deviation is largest near the origin, i.e., within a range of ±100 volts/cm. Relatively marked maxima and minima appear in this region. When the loop is that for a saturated ceramic [Fig. 10(E)] the derived curve is very complicated; see Fig. 12(B). In the region of the small anomalous loop the change of current with potential is extremely rapid.

For portions of the loop lying in quadrants II and IV of Fig. 10(E) σ_d varies only slightly from a constant value except in the neighborhood of the origin and, of course, at maximum field strengths. These results again are to be ascribed to change of domain configuration and redistribution and change of charge between domain walls with change in field. There is a time lag between field and response. The response, in part at least, is probably discontinuous and can be imagined to be a sort of Barkhausen effect as observed for single crystals of $BaTiO_3$ by de Bretteville.[9] The statistical mean of the discontinuities for the many domains appears as a continuous change.

References

1. Nonlinear Dielectric Materials, **E. T. Jaynes.** *Proceedings*, Institute of Radio Engineers, New York, N.Y., vol. 43, 1955, pp. 1733–37.

2. Effect of Grain-Size on the Hysteresis Loop of $BaTiO_3$ Ceramic, **W. Heywang, R. Schöfer.** *Zeitschrift für angewandte Physik*, Berlin, Germany, vol. 8, 1956, pp. 209–13.

3. Depolarization Effects in Polycrystalline $BaTiO_3$, **H. Kniepkamp, W. Heywang.** *Ibid.*, vol. 6, 1954, pp. 385–90.

4. Investigations Concerning Polarization in Barium Titanate Ceramics, **G. W. Marks, D. L. Waidelich, L. A. Monson.** *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 75, Sept. 1956, pp. 469–77.

5. Piezoelectricity (book), **W. G. Cady.** McGraw-Hill Book Company, Inc., New York, N. Y., 1946, chap. 22.

6. Concerning Several Anomalies in the Hysteresis Loops of Ceramic Barium Titanate and Piezolan, **E. Hegenbarth.** *Annalen der Physik*, Leipzig, Germany, vol. 20, 1957, pp. 20–25.

7. The Electric and Optical Behavior of $BaTiO_3$ Single Domain Crystals, **W. J. Merz.** *Physical Review*, New York, N. Y., vol. 76, 1949, pp. 1221–25.

8. Electromagnetics (book), **J. D. Kraus.** McGraw-Hill Book Company, Inc., 1953, chap. 3.

9. Anomalous Effects in Hysteresis Loop of a Single Crystal of $BaTiO_3$, **A. de Bretteville, Jr.** *Physical Review*, vol. 73, 1948, p. 807.

Switching Function Canonical Forms Based on Commutative and Associative Binary Operations

PETER CALINGAERT
NONMEMBER AIEE

The Switching Function

Application of Vector Notation

It is often convenient to consider the arguments of a switching function to be the components zero or 1 of a vector. It is appropriate, therefore, to employ a vector notation for the theory of switching. The notation developed by Iverson[1] for the study of automatic data processing is admirably suited, with a few modifications, to the theory of switching as well. Those features of the Iverson notation that are appropriate have been adapted for the present discussion, and are set forth below.

The dimension $\nu(\mathbf{v})$ of a vector $\mathbf{v}$ is the number of its components $\mathbf{v}_0, \mathbf{v}_1, \ldots, \mathbf{v}_{\nu-1}$. The weight $\sigma(\mathbf{v})$ of a vector is the sum

$$\sigma(\mathbf{v})=\sum_{k=0}^{\nu-1} \mathbf{v}_k$$

of its components. A logical vector is a vector each of whose components is zero or unity. In particular, a vector all of whose components are zero is called the zero vector $\mathbf{0}$; a vector all of whose components are 1 is called the full vector $\boldsymbol{\epsilon}$; a vector whose kth component is 1 and whose remaining components are zero is called the kth unit vector $\boldsymbol{\epsilon}^k$. The dimension of these, as of all other vectors, will be stated specifically only if not clear from the context.

Each member of the set (0,1) is the complement of the other. The complement of a component $\mathbf{v}_k$ of a logical vector is denoted $\bar{\mathbf{v}}_k$. The complement $\bar{\mathbf{v}}$ of a logical vector $\mathbf{v}$ is the vector whose components are the complements of the corresponding components of $\mathbf{v}$. Logical vectors may be partially ordered by the inclusion relation $\geq$. This relation is defined for a pair of vectors of equal dimension as $\mathbf{u}\geq\mathbf{v}$ if and only if $\mathbf{u}_k\geq\mathbf{v}_k$ for all k. Two operations on vectors will now be defined and illustrated, for the sake of clarity, with respect to vectors whose components are not necessarily 0 or 1, although they will later be used only on logical vectors.

The result $\mathbf{y}=\mathbf{u}/\mathbf{v}$ of compressing

Paper 60-1199, recommended by the AIEE Computing Devices Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Chicago, Ill., October 9–14, 1960. Manuscript submitted May 25, 1960; made available for printing August 3, 1960.

Peter Calingaert is with The Computation Laboratory of Harvard University, Cambridge, Mass.

The writer is indebted to Peter G. Neumann, who first pointed out to him the similarity between Pascal's triangle reduced modulo 2 and a chart[2] based on the vector inclusion relation and designed for the detection of monotonicity in switching functions. The advice of Kenneth E. Iverson, Warren L. Semon, and Martin Cohn is gratefully acknowledged.

Table I. Switching Functions of Two Arguments

j	$\mathbf{b}^j$	x_0	x_1	K: 0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0	00	0	0	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1
1	01	0	1	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1
2	10	1	0	0	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1
3	11	1	1	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1

the arbitrary vector $\mathbf{v}$ by the logical vector $\mathbf{u}$ of the same dimension is the vector obtained from $\mathbf{v}$ by suppressing $\mathbf{v}_k$ if and only if $\mathbf{u}_k=0$. It follows that $\nu(\mathbf{y})=\sigma(\mathbf{u})$. For example

$$\mathbf{v}=(S,T,A,R,K) \qquad \nu=5$$

$$\mathbf{u}=(0,1,1,1,0) \qquad \nu=5 \quad \sigma=3$$

$$\mathbf{y}=\mathbf{u}/\mathbf{v}=(T,A,R) \qquad \nu=3$$

The result $\mathbf{w}=\left|\mathbf{t},\mathbf{u},\mathbf{v}\right|$ of meshing the arbitrary vectors $\mathbf{t}$ and $\mathbf{v}$ on the logical vector $\mathbf{u}$ is defined implicitly as the unique vector $\mathbf{w}$ such that $\bar{\mathbf{u}}/\mathbf{w}=\mathbf{t}$ and $\mathbf{u}/\mathbf{w}=\mathbf{v}$. The component $\mathbf{w}_k$ is selected from $\mathbf{t}$ or $\mathbf{v}$ according as $\mathbf{u}_k$ is 0 or 1. The operation is defined only for $\nu(\mathbf{t})=\sigma(\bar{\mathbf{u}})$ and $\nu(\mathbf{v})=\sigma(\mathbf{u})$, and it follows that $\nu(\mathbf{w})=\nu(\mathbf{t})+\nu(\mathbf{v})=\sigma(\bar{\mathbf{u}})+\sigma(\mathbf{u})=\nu(\mathbf{u})$. For example

$$\mathbf{t}=(A,A,A) \qquad \nu=3$$

$$\mathbf{v}=(N,G,R,M) \qquad \nu=4$$

$$\mathbf{u}=(0,1,0,1,1,0,1) \qquad \nu=7 \quad \sigma=4$$

$$\mathbf{w}=\left|\mathbf{t},\mathbf{u},\mathbf{v}\right|=(A,N,A,G,R,A,M) \qquad \nu=7 \|$$

The vector $\mathbf{b}^j$ of dimension ν is the vector whose components are the coefficients of the ν-digit base 2 representation of the nonnegative integer j, taken in natural order, so that $\mathbf{b}_k{}^j$ is the coefficient of $2^{\nu-1-k}$. The vector is defined only for $\nu \geq \log_2(j+1)$. For example, if $\nu=6$, $\mathbf{b}^{26}=(0,1,1,0,1,0)$.

A switching function of n arguments may be defined as a function from the argument set of 2^n vectors $\mathbf{x}$ of n components $\mathbf{x}_0, \ldots, \mathbf{x}_{n-1}$ each 0 or 1, to the value set (0,1). A switching function is uniquely specified by listing its value $F(\mathbf{x})$ for each of the 2^n sets of values, or configurations, of its arguments. The jth configuration of the arguments $\mathbf{x}$ is that for which the argument $\mathbf{x}_k$ is assigned the value $\mathbf{b}_k{}^j$, i.e., $\mathbf{x}=\mathbf{b}^j$. If $\mathbf{f}_j$ is the value $F(\mathbf{b}^j)$ for the jth configuration, then the vector $\mathbf{f}$ of dimension 2^n uniquely specifies the function. The complement $\bar{F}(\mathbf{x})$ of the switching function $F(\mathbf{x})$ is the function specified by the vector $\bar{\mathbf{f}}$. For the jth configuration $\mathbf{x}=\mathbf{b}^j$, the functions $F(\mathbf{x})$ and $\bar{F}(\mathbf{x})$ assume the complementary values $\mathbf{f}_j$ and $\bar{\mathbf{f}}_j$ respectively.

Table II. Commutative and Associative Binary Operations

u	v	K=6 u Δ v	K=8 u ∧ v	K=9 u ∇ v	K=14 u ∨ v
0	0	0	0	1	0
0	1	1	0	0	1
1	0	1	0	0	1
1	1	0	1	1	1

The vectors $\mathbf{f}$ for each of the 16 switching functions of two arguments are listed in Table I. Each is identified by the characteristic number K defined by $\mathbf{f}=\mathbf{b}^K$.

THE CHOICE OF OPERATIONS

To investigate systematically the properties of switching functions it is important to have standard algebraic forms for their representation and manipulation. For maximum utility the representation must be simple and the manipulation easy.

Simplicity of representation is achieved by using algebraic operations requiring few arguments, and by limiting the number of levels over which the operations are iterated. Complementation, the only nontrivial unary operation on the set (0,1), is inadequate to construct an arbitrary switching function from its arguments. Binary operations are therefore required, and it is well known that they are sufficient. No binary operation on the set (0,1) extended on a single level over complemented and uncomplemented arguments is sufficient to represent an arbitrary switching function. Two levels are therefore required and if two different binary operations are used, one on each level, two levels suffice. One binary operation, the primary connective, can be extended over expressions each of which is formed by extending the other binary operation, the secondary connective, over complemented and uncomplemented arguments.

Ease of manipulation is attained by selecting binary operations that are commutative and associative, and such that the secondary connective is distributive over the primary connective. That these properties aid algebraic simplification is shown by the following example in the algebra of real numbers, in which multiplication is distributive over addition.

$$\begin{aligned} &[a(bc)]+[a(db)] \\ &\quad=[a(bc)]+[a(bd)] && \text{by commutativity} \\ &\quad=[(ab)c]+[(ab)d] && \text{by associativity} \\ &\quad=(ab)(c+d) && \text{by distributivity} \end{aligned}$$

The 16 binary operations on the set (0,1) are precisely the 16 switching functions of two arguments listed in Table I. The six functions for $K=0, 3, 5, 10, 12, 15$ are trivial in that their values depend on at most one of the arguments. Of the ten nontrivial functions, those for $K=2, 4, 11, 13$ are asymmetric in the arguments and therefore constitute noncommutative operations. Of the six functions remaining, the associative law does not hold for $K=1$ or for $K=7$. These two operations, the Lukasiewicz and Sheffer functions respectively, are precisely the only two binary operations either of which alone suffices to represent an arbitrary switching function in terms of its arguments. Although the resulting expressions use a single connective, they are extremely cumbersome.

There remain the four functions for $K=6, 8, 9, 14$. It is easily verified that each of the four is a commutative and associative binary operation. They will be denoted by the operation symbols introduced in Table II.

The operation Δ has been called *symmetric difference, sum modulo 2, exclusive or,* and *exclusive disjunction.* The operation ∧ has been called *Boolean product, logical product, and,* and *conjunction.* The operation ∇ has been called *logical equivalence,* and *if and only if.* The operation ∨ has been called *Boolean sum, logical sum, (inclusive) or,* and *(inclusive) disjunction.* A few of the more important properties of the four operations are summarized in Table III.

The binary operations of Table II are extended to vectors of equal dimension as follows. If ⊙ is one of the four operations Δ, ∧, ∇, and ∨, then $\mathbf{u}=\mathbf{v}\odot\mathbf{w}$ if and only if $\mathbf{u}_k=\mathbf{v}_k\odot\mathbf{w}_k$ for all k. The symbols Δ, ⋀, ∇, and ⋁ denote extended application of the operations Δ, ∧, ∇, and ∨ respectively.

CANONICAL FORMS

The four distributive laws stated in Table III are the only ones that hold

Table III. Properties of the Operations

$u \wedge u=u$	$u \vee u=u$	$u \,\Delta\, u=0$	$u \,\nabla\, u=1$
$u \wedge \bar{u}=0$	$u \vee \bar{u}=1$	$u \,\Delta\, \bar{u}=1$	$u \,\nabla\, \bar{u}=0$
$u \wedge 0=0$	$u \vee 0=u$	$u \,\Delta\, 0=u$	$u \,\nabla\, 0=\bar{u}$
$u \wedge 1=u$	$u \vee 1=1$	$u \,\Delta\, 1=\bar{u}$	$u \,\nabla\, 1=u$

$$u \wedge (v \vee w)=(u \wedge v) \vee (u \wedge w) \tag{1}$$

$$u \vee (v \wedge w)=(u \vee v) \wedge (u \vee w) \tag{2}$$

$$u \wedge (v \,\Delta\, w)=(u \wedge v) \,\Delta\, (u \wedge w) \tag{3}$$

$$u \vee (v \,\nabla\, w)=(u \vee v) \,\nabla\, (u \vee w) \tag{4}$$

Table IV. Coefficients, Products, and Sums for a Function of Three Arguments

j	$\mathbf{b}^j$	f	p	g	q	h	r	i	s
0	000	0	$\bar{\mathbf{x}}_0 \wedge \bar{\mathbf{x}}_1 \wedge \bar{\mathbf{x}}_2$	0	$\mathbf{x}_0 \vee \mathbf{x}_1 \vee \mathbf{x}_2$	0	1	0	0
1	001	1	$\bar{\mathbf{x}}_0 \wedge \bar{\mathbf{x}}_1 \wedge \mathbf{x}_2$	1	$\mathbf{x}_0 \vee \mathbf{x}_1 \vee \bar{\mathbf{x}}_2$	1	$\mathbf{x}_2$	0	$\bar{\mathbf{x}}_2$
2	010	1	$\bar{\mathbf{x}}_0 \wedge \mathbf{x}_1 \wedge \bar{\mathbf{x}}_2$	1	$\mathbf{x}_0 \vee \bar{\mathbf{x}}_1 \vee \mathbf{x}_2$	1	$\mathbf{x}_1$	0	$\bar{\mathbf{x}}_1$
3	011	0	$\bar{\mathbf{x}}_0 \wedge \mathbf{x}_1 \wedge \mathbf{x}_2$	0	$\mathbf{x}_0 \vee \bar{\mathbf{x}}_1 \vee \bar{\mathbf{x}}_2$	0	$\mathbf{x}_1 \wedge \mathbf{x}_2$	1	$\bar{\mathbf{x}}_1 \vee \bar{\mathbf{x}}_2$
4	100	1	$\mathbf{x}_0 \wedge \bar{\mathbf{x}}_1 \wedge \bar{\mathbf{x}}_2$	1	$\bar{\mathbf{x}}_0 \vee \mathbf{x}_1 \vee \mathbf{x}_2$	1	$\mathbf{x}_0$	0	$\bar{\mathbf{x}}_0$
5	101	0	$\mathbf{x}_0 \wedge \bar{\mathbf{x}}_1 \wedge \mathbf{x}_2$	0	$\bar{\mathbf{x}}_0 \vee \mathbf{x}_1 \vee \bar{\mathbf{x}}_2$	0	$\mathbf{x}_0 \wedge \mathbf{x}_2$	1	$\bar{\mathbf{x}}_0 \vee \bar{\mathbf{x}}_2$
6	110	0	$\mathbf{x}_0 \wedge \mathbf{x}_1 \wedge \bar{\mathbf{x}}_2$	0	$\bar{\mathbf{x}}_0 \vee \bar{\mathbf{x}}_1 \vee \mathbf{x}_2$	0	$\mathbf{x}_0 \wedge \mathbf{x}_1$	1	$\bar{\mathbf{x}}_0 \vee \bar{\mathbf{x}}_1$
7	111	1	$\mathbf{x}_0 \wedge \mathbf{x}_1 \wedge \mathbf{x}_2$	1	$\bar{\mathbf{x}}_0 \vee \bar{\mathbf{x}}_1 \vee \bar{\mathbf{x}}_2$	0	$\mathbf{x}_0 \wedge \mathbf{x}_1 \wedge \mathbf{x}_2$	1	$\bar{\mathbf{x}}_0 \vee \bar{\mathbf{x}}_1 \vee \bar{\mathbf{x}}_2$

among the four commutative and associative operations. The operations in each distributive law will be used as the connectives of a standard, or canonical, form of an arbitrary switching function $F(\mathbf{x})$ of n arguments.

$$u\wedge(v\vee w)=(u\wedge v)\vee(u\wedge w) \tag{1}$$

$$u\vee(v\wedge w)=(u\vee v)\wedge(u\vee w) \tag{2}$$

$$u\wedge(v\Delta w)=(u\wedge v)\Delta(u\wedge w) \tag{3}$$

$$u\vee(v\nabla w)=(u\vee v)\nabla(u\vee w) \tag{4}$$

Corresponding to equation 1 is a form with primary connective $\vee$ and secondary connective $\wedge$, defined by

$$\mathbf{p}_j=\bigwedge_{k=0}^{n-1}(\mathbf{x}_k \nabla \mathbf{b}_k{}^j)$$

$$F(\mathbf{x})=\bigvee_{j=0}^{2^n-1}(\mathbf{f}_j \wedge \mathbf{p}_j) \tag{5}$$

The products $\mathbf{p}_j$ have been called fundamental products, constituents of unity, and minterms. The canonical form, equation 5, has been called the sum of products canonical form and the disjunctive canonical form.

Corresponding to equation 2 is a form with primary connective $\wedge$ and secondary connective $\vee$, defined by

$$\mathbf{q}_j=\bigvee_{k=0}^{n-1}(\mathbf{x}_k \Delta \mathbf{b}_k{}^j)$$

$$F(\mathbf{x})=\bigwedge_{j=0}^{2^n-1}(\mathbf{g}_j \vee \mathbf{q}_j) \tag{6}$$

The sums $\mathbf{q}_j$ have been called fundamental sums, constituents of zero, and maxterms. The canonical form, equation 6, has been called the product of sums canonical form and the conjunctive canonical form.

Corresponding to equation 3 is a form with primary connective Δ and secondary connective $\wedge$, defined by

$$\mathbf{r}_j=\bigwedge_{k=0}^{n-1}(\mathbf{x}_k \vee \mathbf{b}_k{}^j)$$

$$F(\mathbf{x})=\mathop{\Delta}_{j=0}^{2^n-1}(\mathbf{h}_j \wedge \mathbf{r}_j) \tag{7}$$

The products $\mathbf{r}_j$ may be called reduced fundamental products, constituents of unity, or minterms. The canonical form, equation 7, may be called the delta canonical form or the exclusive disjunctive canonical form.

Corresponding to equation 4 is a form with primary connective ∇ and secondary connective $\vee$, defined by

$$\mathbf{s}_j=\bigvee_{k=0}^{n-1}(\bar{\mathbf{x}}_k \wedge \mathbf{b}_k{}^j)$$

$$F(\mathbf{x})=\mathop{\nabla}_{j=0}^{2^n-1}(\mathbf{i}_j \vee \mathbf{s}_j) \tag{8}$$

The sums $\mathbf{s}_j$ may be called reduced fundamental sums, constituents of zero, or maxterms. The canonical form, equation 8, is believed to be new and may be called the del canonical form or the equivalence canonical form.

The logical vectors **f**, **g**, **h**, and **i** are determined by the switching function $F(\mathbf{x})$. For example, let $F(\mathbf{x})$ be the function of three arguments, which assumes the value unity if and only if exactly one or exactly three of the arguments are unity. The coefficients **f**, **g**, **h**, and **i**, the products **p** and **r**, and the sums **q** and **s** are all listed for this function in Table IV.

The connective joining $\mathbf{x}_k$ and $\mathbf{b}_k{}^j$ in the definition of $\mathbf{p}_j$, $\mathbf{q}_j$, $\mathbf{r}_j$, or $\mathbf{s}_j$ serves only to help specify whether $\mathbf{x}_k$ is to be included or not, or complemented or not, in the product or sum being defined. Contrary to appearances, it does not constitute a third connective in any of the canonical forms, as may be verified by inspection of Table IV.

Writing out the canonical forms, equations 5 through 8, for the example yields, upon replacement of the sign $\wedge$ by juxtaposition and after some elementary algebraic simplification, the following four expressions:

$$F(\mathbf{x})=\bar{\mathbf{x}}_0\bar{\mathbf{x}}_1\mathbf{x}_2 \vee \bar{\mathbf{x}}_0\mathbf{x}_1\bar{\mathbf{x}}_2 \vee \mathbf{x}_0\bar{\mathbf{x}}_1\bar{\mathbf{x}}_2 \vee \mathbf{x}_0\mathbf{x}_1\mathbf{x}_2$$

$$F(\mathbf{x})=(\mathbf{x}_0 \vee \mathbf{x}_1 \vee \mathbf{x}_2)(\mathbf{x}_0 \vee \bar{\mathbf{x}}_1 \vee \bar{\mathbf{x}}_2)(\bar{\mathbf{x}}_0 \vee \mathbf{x}_1 \vee \bar{\mathbf{x}}_2)(\bar{\mathbf{x}}_0 \vee \bar{\mathbf{x}}_1 \vee \mathbf{x}_2)$$

$$F(\mathbf{x})=\mathbf{x}_2 \Delta \mathbf{x}_1 \Delta \mathbf{x}_0$$

$$F(\mathbf{x})=0 \nabla \bar{\mathbf{x}}_2 \nabla \bar{\mathbf{x}}_1 \nabla \bar{\mathbf{x}}_0$$

It is evident from the example that some canonical forms yield expressions more concise than others for a given function. Some may be better suited than others to describing switching circuits in a given technology. It is obviously necessary to have means of translating from one representation to another. For any particular translation, straightforward algebraic manipulation can be employed. For example, the disjunctive form can be translated to the delta form in three steps. The first is the replacement of $\vee$ by Δ, which leaves the function unchanged because the individual products are disjoint. The second is the replacement of $\bar{\mathbf{x}}_i$ by its equivalent $(\mathbf{x}_i \Delta 1)$. The third is the application of the properties listed in Table III, together with the commutativity and associativity of $\vee$ and Δ, to reduce the resulting expression to the form of equation 7. For the function of Table IV these steps yield

$$\begin{aligned}F(\mathbf{x})&=\bar{\mathbf{x}}_0\bar{\mathbf{x}}_1\mathbf{x}_2 \Delta \bar{\mathbf{x}}_0\mathbf{x}_1\bar{\mathbf{x}}_2 \Delta \mathbf{x}_0\bar{\mathbf{x}}_1\bar{\mathbf{x}}_2 \Delta \mathbf{x}_0\mathbf{x}_1\mathbf{x}_2\\&=(\mathbf{x}_0 \Delta 1)(\mathbf{x}_1 \Delta 1)\mathbf{x}_2 \Delta (\mathbf{x}_0 \Delta 1)\mathbf{x}_1(\mathbf{x}_2 \Delta 1)\\&\qquad \Delta \mathbf{x}_0(\mathbf{x}_1 \Delta 1)(\mathbf{x}_2 \Delta 1) \Delta \mathbf{x}_0\mathbf{x}_1\mathbf{x}_2\\&=\mathbf{x}_0\mathbf{x}_1\mathbf{x}_2 \Delta \mathbf{x}_0\mathbf{x}_2 \Delta \mathbf{x}_1\mathbf{x}_2 \Delta \mathbf{x}_2 \Delta \mathbf{x}_0\mathbf{x}_1\mathbf{x}_2\\&\qquad \Delta \mathbf{x}_0\mathbf{x}_1 \Delta \mathbf{x}_1\mathbf{x}_2 \Delta \mathbf{x}_1 \Delta \mathbf{x}_0\mathbf{x}_1\mathbf{x}_2 \Delta \mathbf{x}_0\mathbf{x}_1\\&\qquad \Delta \mathbf{x}_0\mathbf{x}_2 \Delta \mathbf{x}_0 \Delta \mathbf{x}_0\mathbf{x}_1\mathbf{x}_2\\&=\mathbf{x}_0 \Delta \mathbf{x}_1 \Delta \mathbf{x}_2\end{aligned}$$

Such tedious manipulation can be avoided by the use of the transformations that carry the coefficient vector of one representation into the coefficient vector of another. It will be shown, for example, that the translation just performed algebraically can be effected by the transformation

$$\left\|\begin{matrix}1&0&0&0&0&0&0&0\\1&1&0&0&0&0&0&0\\1&0&1&0&0&0&0&0\\1&1&1&1&0&0&0&0\\1&0&0&0&1&0&0&0\\1&1&0&0&1&1&0&0\\1&0&1&0&1&0&1&0\\1&1&1&1&1&1&1&1\end{matrix}\right\| \left\|\begin{matrix}0\\1\\1\\0\\1\\0\\0\\1\end{matrix}\right\| = \left\|\begin{matrix}0\\1\\1\\0\\1\\0\\0\\0\end{matrix}\right\|$$

which may be written as $\mathbf{h}=\mathbf{A}\mathbf{f}$, where $\mathbf{A}$ is the matrix shown and where matrix multiplication is performed using modulo 2 arithmetic. The desired transformations will all be determined explicitly with the aid of two further results. These are a theorem on the partial ordering of Boolean vectors, and a derivation of the delta canonical form by means of the partial difference operation.

Partial Ordering of Boolean Vectors

PARITY OF BINOMIAL COEFFICIENTS

It is assumed throughout this discussion that arguments of binomial coefficients are restricted to nonnegative integral values. Use will be made of two relations on binomial coefficients, stated here without proof.

$$\begin{bmatrix}x+y\\z\end{bmatrix}=\sum_{h=0}^{z}\begin{bmatrix}x\\z-h\end{bmatrix}\begin{bmatrix}y\\h\end{bmatrix} \tag{9}$$

$$\begin{bmatrix} y \\ x \end{bmatrix} = \sum_{h=0}^{i} \begin{bmatrix} i \\ h \end{bmatrix} \begin{bmatrix} y-i \\ x-h \end{bmatrix} \tag{10}$$

for $0 \leq i \leq y$.

The following lemmas concern the parity of binomial coefficients. It is assumed in each lemma that n is integral.

Lemma 1

$$\begin{bmatrix} 2^n \\ k \end{bmatrix} \equiv 1 \pmod 2 \longleftrightarrow k=0 \text{ or } k=2^n$$

Proof: Sufficiency for all k and necessity for $k=0$ and $k=2^n$ follow from the relations

$$\begin{bmatrix} 2^n \\ 0 \end{bmatrix} = 1 \text{ and } \begin{bmatrix} 2^n \\ 2^n \end{bmatrix} = 1$$

Since

$$\begin{bmatrix} 2^n \\ k \end{bmatrix} = 0$$

for $k > 2^n$, necessity need be further demonstrated only for $0<k<2^n$. The demonstration proceeds by induction on n. The proposition is trivially true for $n=0$ and, since

$$\begin{bmatrix} 2 \\ 1 \end{bmatrix} = 2$$

it is also true for $n=1$. Let it be true by hypothesis for $n-1$. In considering the coefficient

$$\begin{bmatrix} 2^n \\ k \end{bmatrix}$$

no generality is lost by assuming that $k \leq 2^{n-1}$, since

$$\begin{bmatrix} 2^n \\ k \end{bmatrix} = \begin{bmatrix} 2^n \\ 2^n-k \end{bmatrix}$$

By equation 9

$$\begin{bmatrix} 2^n \\ k \end{bmatrix} = \begin{bmatrix} 2^{n-1}+2^{n-1} \\ k \end{bmatrix} = \sum_{j=0}^{k} \begin{bmatrix} 2^{n-1} \\ k-j \end{bmatrix} \begin{bmatrix} 2^{n-1} \\ j \end{bmatrix}$$

If $k<2^{n-1}$, the factor

$$\begin{bmatrix} 2^{n-1} \\ k-j \end{bmatrix}$$

is even, by the inductive hypothesis, for $j \neq k$, and the factor

$$\begin{bmatrix} 2^{n-1} \\ j \end{bmatrix}$$

is likewise even for $j \neq 0$. Since $k \neq 0$, at least one factor of each term is even and so, therefore, is their sum. If $k=2^{n-1}$, the sum reduces to

$$\sum_{j=0}^{2^{n-1}} \begin{bmatrix} 2^{n-1} \\ 2^{n-1}-j \end{bmatrix} \begin{bmatrix} 2^{n-1} \\ j \end{bmatrix} = \sum_{j=0}^{2^{n-1}} \begin{bmatrix} 2^{n-1} \\ j \end{bmatrix}^2$$

Since the first and last terms of this sum are unity and by hypothesis all other terms are even, the sum is even.

Lemma 2

$$\left.\begin{matrix} a \leq b \\ a < 2^n \end{matrix}\right\} \rightarrow \begin{bmatrix} b+2^n \\ a \end{bmatrix} \equiv \begin{bmatrix} b \\ a \end{bmatrix} \pmod 2$$

Proof: By equation 9

$$\begin{bmatrix} b+2^n \\ a \end{bmatrix} = \sum_{k=0}^{a} \begin{bmatrix} b \\ a-k \end{bmatrix} \begin{bmatrix} 2^n \\ k \end{bmatrix}$$

Since, by lemma 1, the second factor

$$\begin{bmatrix} 2^n \\ k \end{bmatrix}$$

is odd if and only if $k=0$, the sum will be odd if and only if the first factor

$$\begin{bmatrix} b \\ a-0 \end{bmatrix}$$

is odd. The lemma follows immediately.

Lemma 3

$$\left.\begin{matrix} b < 2^n \\ a < 2^n \end{matrix}\right\} \rightarrow \begin{bmatrix} b+2^n \\ a+2^n \end{bmatrix} \equiv \begin{bmatrix} b \\ a \end{bmatrix} \pmod 2$$

Proof: By equation 10

$$\begin{bmatrix} b+2^n \\ a+2^n \end{bmatrix} = \sum_{k=0}^{2^n} \begin{bmatrix} 2^n \\ k \end{bmatrix} \begin{bmatrix} b \\ a+2^n-k \end{bmatrix}$$

Since, by lemma 1, the factor

$$\begin{bmatrix} 2^n \\ k \end{bmatrix}$$

is odd if and only if $k=0$ or $k=2^n$, only for these two values of k may there be an odd term in the sum. Hence

$$\begin{bmatrix} b+2^n \\ a+2^n \end{bmatrix} \equiv \begin{bmatrix} b \\ a+2^n \end{bmatrix} + \begin{bmatrix} b \\ a \end{bmatrix} \pmod 2$$

Since $a+2^n > b$

$$\begin{bmatrix} b \\ a+2^n \end{bmatrix} = 0$$

and the lemma follows immediately.

Partial Ordering and Parity

Let $\mathbf{b}^u$ and $\mathbf{b}^v$ be the logical vectors of dimension n whose components are the coefficients of the base 2 representations of the integers u and v, where $u<2^n$ and $v<2^n$. The main result relating the partial ordering of logical vectors to the parity of binomial coefficients is established in the following theorem:

Theorem 1

$$\mathbf{b}^v \geq \mathbf{b}^u \longleftrightarrow \begin{bmatrix} v \\ u \end{bmatrix} \equiv 1 \pmod 2$$

Proof: The theorem is easily verified for $n=1$, and the proof proceeds by induction on n. The equations

$$u = 2^{n-1}\mathbf{b}_0^u + u'$$

$$v = 2^{n-1}\mathbf{b}_0^v + v'$$

uniquely determine the integers u' and v' represented by the $n-1$ rightmost components of $\mathbf{b}^u$ and $\mathbf{b}^v$ respectively. The corresponding logical vectors of dimension $n-1$ are $\mathbf{b}^{u'}$ and $\mathbf{b}^{v'}$.

Case 1: If $\mathbf{b}_0^v \geq \mathbf{b}_0^u$ does not hold, then $\mathbf{b}^v \geq \mathbf{b}^u$ does not hold, and since $v<u$ in this case

$$\begin{bmatrix} v \\ u \end{bmatrix} = 0$$

and the theorem is verified.

Case 2: If $\mathbf{b}_0^v \geq \mathbf{b}_0^u$, then

$$\mathbf{b}^v \geq \mathbf{b}^u \longleftrightarrow \mathbf{b}^{v'} \geq \mathbf{b}^{u'} \tag{11}$$

Moreover,

$$\begin{bmatrix} v \\ u \end{bmatrix} = \begin{bmatrix} v'+2^{n-1}\mathbf{b}_0^v \\ u'+2^{n-1}\mathbf{b}_0^u \end{bmatrix} \equiv \begin{bmatrix} v' \\ u' \end{bmatrix} \pmod 2$$

identically for

$$\mathbf{b}_0^v = \mathbf{b}_0^u = 0$$

and by lemma 2 for

$$\mathbf{b}_0^v = 1,\ \mathbf{b}_0^u = 0$$

and by lemma 3 for

$$\mathbf{b}_0^v = \mathbf{b}_0^u = 1$$

Hence

$$\begin{bmatrix} v \\ u \end{bmatrix} \equiv 1 \pmod 2 \longleftrightarrow \begin{bmatrix} v' \\ u' \end{bmatrix} \equiv 1 \pmod 2 \tag{12}$$

But, by the inductive hypothesis,

$$\mathbf{b}^{v'} \geq \mathbf{b}^{u'} \longleftrightarrow \begin{bmatrix} v' \\ u' \end{bmatrix} \equiv 1 \pmod 2$$

Hence, from equations 11 and 12,

$$\mathbf{b}^v \geq \mathbf{b}^u \longleftrightarrow \begin{bmatrix} v \\ u \end{bmatrix} \equiv 1 \pmod 2$$

and the theorem is established.

An Alternative Proof

It has been called to the writer's attention that theorem 1 can be deduced from a theorem of Kummer.[3] Consider the prime radix p representations of the integers a and b:

$$a = \mathbf{a}_0 p^{n-1} + \ldots + \mathbf{a}_{n-1} p^0$$

$$b = \mathbf{b}_0 p^{n-1} + \ldots + \mathbf{b}_{n-1} p^0$$

The addition of a and b defines the sum $d=a+b$ and the logical vector $\mathbf{c}$ whose kth component $\mathbf{c}_k$ is the carry generated in the kth column. Kummer proved that the highest power p^N of p that divides

$$\begin{bmatrix} a+b \\ a \end{bmatrix}$$

is given by $N=\sigma(\mathbf{c})$.

Setting $a+b=v$, $a=u$, and $p=2$, it is easily shown that $\mathbf{b}^v \geq \mathbf{b}^u$ if and only if $\mathbf{c}=\mathbf{0}$ in the addition of a to b, i.e., of u to $v-u$. The sufficiency is obvious. The necessity follows from the observation that there exists no carry into the least significant column. This will be indicated, by extension, as $\mathbf{c}_n=0$. It should be observed that if $\mathbf{b}^v \geq \mathbf{b}^u$, then $\nu(\mathbf{d}) = n$ and $\mathbf{d}_j$ is the sum digit in the same column as $\mathbf{a}_j$ and $\mathbf{b}_j$. If $\mathbf{c}_i=1$ for some column i, then there exists at least one column j, $i \leq j < n$, such that $\mathbf{c}_j=1$ and $\mathbf{c}_{j+1}=0$. It follows that $\mathbf{a}_j=\mathbf{b}_j=1$ and $\mathbf{d}_j=0$. But $\mathbf{d}_j=\mathbf{b}_j{}^v$ and $\mathbf{a}_j=\mathbf{b}_j{}^u$ and it follows from the fact that $\mathbf{b}_j{}^v \geq \mathbf{b}_j{}^u$ does not hold, that $\mathbf{b}^v \geq \mathbf{b}^u$ does not hold either, which is contrary to hypothesis. Hence $\mathbf{c}_i=0$ for all i, $\mathbf{c}=\mathbf{0}$, and $N=\sigma(\mathbf{c})=\sigma(\mathbf{0})=0$ if and only if $\mathbf{b}^v \geq \mathbf{b}^u$. Since 2^1 divides

$$\begin{bmatrix} v \\ u \end{bmatrix} \text{ if and only if } N>0$$

theorem 1 follows.

The Partial Difference Operation

Following Reed[4] and Akers,[5] the partial difference of the switching function $F(\mathbf{x})$ with respect to the argument $\mathbf{x}_k$ is the symmetric difference of the two function values $F(\bar{\mathbf{e}}^k \wedge \mathbf{x})$ and $F(\mathbf{e}^k \vee \mathbf{x})$ obtained by setting $\mathbf{x}_k=0$ and $\mathbf{x}_k=1$ respectively, in $F(\mathbf{x})$. The partial difference of $F(\mathbf{x})$ with respect to $m>1$ arguments is the extended symmetric difference (unique by the commutativity and associativity of Δ) of the 2^m function values obtained by assigning all configurations of values to those m arguments in $F(\mathbf{x})$.

The partial difference operation can be defined formally with the aid of the identity

$$\mathbf{x}=|\bar{\mathbf{u}}/\mathbf{x},\mathbf{u},\mathbf{u}/\mathbf{x}|$$

where $\mathbf{u}$ is any logical vector of the same dimension as $\mathbf{x}$. The vector resulting from the substitution in $\mathbf{x}$ of the coefficients of the m-digit base 2 representation of the integer k for the m arguments corresponding to components 1 of $\mathbf{u}$ is clearly

$$|\bar{\mathbf{u}}/\mathbf{x},\mathbf{u},\mathbf{b}^k|$$

where $\nu(\mathbf{b}^k)=\sigma(\mathbf{u})=m$.

The partial difference

$$\overset{\partial}{\mathbf{u}}F(\mathbf{x})$$

of $F(\mathbf{x})$ with respect to the arguments specified by $\mathbf{u}$ is defined as

$$\overset{\partial}{\mathbf{u}}F(\mathbf{x})=\underset{k=0}{\overset{2^{\sigma(\mathbf{u})}-1}{\Delta}} F(|\bar{\mathbf{u}}/\mathbf{x},\mathbf{u},\mathbf{b}^k|)$$

For example, let $n=4$ and $\mathbf{u}=(0,1,0,1)$. Then $\sigma(\mathbf{u})=2$ and

$$\begin{aligned}\overset{\partial}{\mathbf{u}}F(\mathbf{x})&=\underset{k=0}{\overset{3}{\Delta}} F(|(\mathbf{x}_0,\mathbf{x}_2),\mathbf{u},\mathbf{b}^k|)\\ &=F(\mathbf{x}_0,0,\mathbf{x}_2,0)\ \Delta\ F(\mathbf{x}_0,0,\mathbf{x}_2,1)\ \Delta\\ &\qquad F(\mathbf{x}_0,1,\mathbf{x}_2,0)\ \Delta\ F(\mathbf{x}_0,1,\mathbf{x}_2,1)\end{aligned}$$

The three special cases $\mathbf{u}=\mathbf{0}$, $\mathbf{u}=\mathbf{e}^k$, and $\mathbf{u}=\mathbf{e}$ yield

$$\overset{\partial}{\mathbf{0}}F(\mathbf{x})=F(\mathbf{x})$$

$$\overset{\partial}{\mathbf{e}^k}F(\mathbf{x})=F(\bar{\mathbf{e}}^k \wedge \mathbf{x})\ \Delta\ F(\mathbf{e}^k \vee \mathbf{x})$$

$$\overset{\partial}{\mathbf{e}}F(\mathbf{x})=\underset{j=0}{\overset{2^n-1}{\Delta}} F(\mathbf{b}^j)=\underset{j=0}{\overset{2^n-1}{\Delta}} \mathbf{f}_j$$

The partial difference operation can be used to expand an arbitrary switching function about its arguments, in a form with Δ as the primary connective. The expansion about $\mathbf{x}_k$ is derived as follows:

$$\begin{aligned}F(\mathbf{x})&=\bar{\mathbf{x}}_k F(\bar{\mathbf{e}}^k \wedge \mathbf{x}) \vee \mathbf{x}_k F(\mathbf{e}^k \vee \mathbf{x})\\ &=\bar{\mathbf{x}}_k F(\bar{\mathbf{e}}^k \wedge \mathbf{x})\ \Delta\ \mathbf{x}_k F(\mathbf{e}^k \vee \mathbf{x})\\ &=(1\ \Delta\ \mathbf{x}_k)F(\bar{\mathbf{e}}^k \wedge \mathbf{x})\ \Delta\ \mathbf{x}_k F(\mathbf{e}^k \vee \mathbf{x})\\ &=F(\bar{\mathbf{e}}^k \wedge \mathbf{x})\ \Delta\ \mathbf{x}_k[F(\bar{\mathbf{e}}^k \wedge \mathbf{x})\ \Delta\ F(\mathbf{e}^k \vee \mathbf{x})]\\ &=F(\bar{\mathbf{e}}^k \wedge \mathbf{x})\ \Delta\ \mathbf{x}_k \overset{\partial}{\mathbf{e}^k}F(\mathbf{x})\end{aligned}$$

If this expansion is performed about each argument in succession, an expression is obtained for $F(\mathbf{x})$ in terms of the products $\mathbf{r}$, defined earlier, and of the partial differences of $F(\mathbf{x})$ with respect to the variables of these products. In particular, it can be shown[5] that

$$F(\mathbf{x})=\underset{j=0}{\overset{2^n-1}{\Delta}}\left[\mathbf{r}_j \wedge \overset{\partial}{\mathbf{b}^j}F(\mathbf{b}^j \wedge \mathbf{x})\right] \tag{13}$$

In each partial difference of equation 13, all arguments have been set equal to zero except those with respect to which the partial difference is taken. For $n=2$, for example:

$$\begin{aligned}F(\mathbf{x}_0,\mathbf{x}_1)&=\overset{\partial}{\mathbf{b}^0}F(\mathbf{b}^0 \wedge \mathbf{x})\ \Delta\ \mathbf{x}_1 \overset{\partial}{\mathbf{b}^1}F(\mathbf{b}^1 \wedge \mathbf{x})\ \Delta\\ &\qquad \mathbf{x}_0 \overset{\partial}{\mathbf{b}^2}F(\mathbf{b}^2 \wedge \mathbf{x})\ \Delta\ \mathbf{x}_0\mathbf{x}_1 \overset{\partial}{\mathbf{b}^3}F(\mathbf{b}^3 \wedge \mathbf{x})\\ &=\overset{\partial}{\mathbf{0}}F(0,0)\ \Delta\ \mathbf{x}_1 \overset{\partial}{\mathbf{b}^1}F(0,\mathbf{x}_1)\ \Delta\\ &\qquad \mathbf{x}_0 \overset{\partial}{\mathbf{b}^2}F(\mathbf{x}_0,0)\ \Delta\ \mathbf{x}_0\mathbf{x}_1 \overset{\partial}{\mathbf{e}}F(\mathbf{x}_0,\mathbf{x}_1)\\ &=F(0,0)\ \Delta\ \mathbf{x}_1[F(0,0)\ \Delta\ F(0,1)]\ \Delta\\ &\qquad \mathbf{x}_0[F(0,0)\ \Delta\ F(1,0)]\ \Delta\ \mathbf{x}_0\mathbf{x}_1\\ &\qquad [F(0,0)\ \Delta\ F(0,1)\ \Delta\ F(1,0)\ \Delta\\ &\qquad F(1,1)]\end{aligned}$$

Transformations Among the Canonical Forms

CALCULATION OF **h** FROM **f**

Comparison of equation 13 with equation 7 provides the identification of the coefficient $\mathbf{h}_j$ as

$$\mathbf{h}_j=\overset{\partial}{\mathbf{b}^j}F(\mathbf{b}^j \wedge \mathbf{x})$$

$$\mathbf{h}_j=\overset{\partial}{\mathbf{b}^j}F(|\mathbf{0},\mathbf{b}^j,\mathbf{b}^j/\mathbf{x}|)$$

$$\mathbf{h}_j=\underset{k=0}{\overset{2^{\sigma(\mathbf{b}^j)}-1}{\Delta}} F(|\mathbf{0},\mathbf{b}^j,\mathbf{b}^k|) \tag{14}$$

where $\nu(\mathbf{b}^k)=\sigma(\mathbf{b}^j)$ and $\nu(\mathbf{b}^j)=n$. The term for $k=2^{\sigma(\mathbf{b}^j)}-1=2^{\nu(\mathbf{b}^k)}-1$ is $F(|\mathbf{0},\mathbf{b}^j,\mathbf{e}|)=F(\mathbf{b}^j)$. This term is obtained from $F(\mathbf{x})$ by assigning the value 1 to all arguments specified by $\mathbf{b}^j$ and the value zero to all others. The term for any other value of k is that resulting from the assignment of some other configuration to the arguments specified by $\mathbf{b}^j$. Each such assignment will yield a vector included by $\mathbf{b}^j$. Equation 14 may therefore be rewritten as

$$\mathbf{h}_j=\underset{k \in S_j}{\Delta} F(\mathbf{b}^k)$$

where $\nu(\mathbf{b}^k)$ is now n and where $S_j=\{k|\mathbf{b}^j \geq \mathbf{b}^k\}$ is the set of integers whose vector representations are included by the vector $\mathbf{b}^j$. The dimension of the set S_j is clearly $2^{\sigma(\mathbf{b}^j)}$.

But, by theorem 1

$$\mathbf{b}^j \geq \mathbf{b}^k \longleftrightarrow \begin{bmatrix} j \\ k \end{bmatrix}_2=1$$

where

$$\begin{bmatrix} j \\ k \end{bmatrix}_2$$ is the residue modulo 2 of the binomial coefficient $$\begin{bmatrix} j \\ k \end{bmatrix}$$

It follows that

$$\mathbf{h}_j=\underset{k=0}{\overset{j}{\Delta}} \mathbf{f}_k \begin{bmatrix} j \\ k \end{bmatrix}_2$$

and, since

$$\begin{bmatrix} j \\ k \end{bmatrix}=0 \text{ for } k>j$$

it follows further that

$$\mathbf{h}_j=\underset{k=0}{\overset{2^n-1}{\Delta}} \mathbf{f}_k \begin{bmatrix} j \\ k \end{bmatrix}_2$$

The transformation of the coefficients $\mathbf{f}$ into the coefficients $\mathbf{h}$ can therefore be expressed as

$$\mathbf{h}=\mathbf{A}\mathbf{f} \tag{15}$$

where matrix multiplication is defined with the operations Δ and $\wedge$, and $\mathbf{A}$ is the square matrix of dimension $r=2^n$ whose general element is

$$\mathbf{A}^{j}_{k}=\begin{bmatrix} j \\ k \end{bmatrix}_2$$

where $j=0, \ldots, r-1$ and $k=0, \ldots, r-1$ are the row and column indexes respectively. The matrices **A** for $n=1, 2, 3$ are

$$\left\|\begin{array}{c|c} 1 & 0 \\ \hline 1 & 1 \end{array}\right\| \quad \left\|\begin{array}{cc|cc} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ \hline 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 \end{array}\right\| \quad \left\|\begin{array}{cccc|cccc} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ \hline 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{array}\right\|$$

The matrices have been partitioned to emphasize their recursive construction.

CALCULATION OF f FROM h

Muller[6] has stated that for $r=2^n$ the transformation $\mathbf{h}=\mathbf{Af}$ is its own inverse. This can be proved in general, and the restriction that the dimension of **A** be a power of 2 is not necessary. The expression $\mathbf{h}=\mathbf{Af}$ represents a system of 2^n linear equations in 2^n unknowns in the field J_2 of integers modulo 2. The system possesses a unique solution if and only if the determinant δ of **A** is not zero. Since the matrix **A** is triangular

$$\delta=\prod_{j=0}^{r-1}\mathbf{A}^{j}_{j}=\prod_{j=0}^{r-1}1=1$$

The system of equation 15 therefore possesses a unique inverse for vectors of any nonzero dimension. It is readily shown that the inverse of **A** is **A** itself. This can be demonstrated by induction for $r=2^n$, using the recursiveness noted above. Since it is true for all $r>0$, a more general proof is of interest.

Theorem 2

If **A** is a square matrix of dimension r with general element

$$\mathbf{A}^{j}_{k}=\begin{bmatrix} j \\ k \end{bmatrix}_2$$

in the field J_2, the product $\mathbf{B}=\mathbf{AA}$ is the identity matrix.

Proof:

$$\mathbf{B}^{j}_{k}=\mathop{\Delta}_{l=0}^{r-1}\mathbf{A}^{j}_{l}\mathbf{A}^{l}_{k}=\mathop{\Delta}_{l=0}^{r-1}\begin{bmatrix} j \\ l \end{bmatrix}_2\begin{bmatrix} l \\ k \end{bmatrix}_2$$

Case 1: $j=k$

$$\mathbf{B}^{j}_{k}=\mathop{\Delta}_{l=0}^{r-1}\begin{bmatrix} j \\ l \end{bmatrix}_2\begin{bmatrix} l \\ j \end{bmatrix}_2$$

But

$$\begin{bmatrix} j \\ l \end{bmatrix}_2\begin{bmatrix} l \\ j \end{bmatrix}_2=1$$

if and only if $l=j$. Hence

$$\mathbf{B}^{j}_{k}=1$$

Case 2: $j<k$

Either $l>j$ or $l<k$ (or both). If $l>j$, the first factor of the lth product is zero; if $l<k$, the second factor of the lth product is zero. Since each product is zero, so is their sum. Hence

$$\mathbf{B}^{j}_{k}=0$$

Case 3: $j>k$

If $l>j$, the first factor of the lth product is zero; if $l<k$, the second factor of the lth product is zero. The only nonzero contribution to

$$\mathbf{B}^{j}_{k}$$

is from products for which $k\leq l\leq j$, and

$$\mathbf{B}^{j}_{k}=\mathop{\Delta}_{l=k}^{j}\begin{bmatrix} j \\ l \end{bmatrix}_2\begin{bmatrix} l \\ k \end{bmatrix}_2$$

But

$$\begin{bmatrix} j \\ l \end{bmatrix}_2\begin{bmatrix} l \\ k \end{bmatrix}_2=1\longleftrightarrow\begin{bmatrix} j \\ l \end{bmatrix}_2=\begin{bmatrix} l \\ k \end{bmatrix}_2=1$$

Therefore by theorem 1

$$\begin{bmatrix} j \\ l \end{bmatrix}_2\begin{bmatrix} l \\ k \end{bmatrix}_2=1\longleftrightarrow\mathbf{b}^j\geq\mathbf{b}^l \text{ and } \mathbf{b}^l\geq\mathbf{b}^k$$

Then either all products are zero or there exists an integer l such that $\mathbf{b}^j\geq\mathbf{b}^l\geq\mathbf{b}^k$. In the latter event $\mathbf{b}^j$ and $\mathbf{b}^k$ differ in $q>0$ components, because $j\neq k$. Each of the 2^q assignments of values to those components yields a unique vector $\mathbf{b}^l$ such that $k\leq l\leq j$ and

$$\begin{bmatrix} j \\ l \end{bmatrix}_2\begin{bmatrix} l \\ k \end{bmatrix}_2=1$$

Then exactly 2^q products equal 1 and their modulo 2 sum

$$\mathbf{B}^{j}_{k}=0$$

Hence the theorem is proved.

Since **A** is self-inverse, it follows immediately from $\mathbf{h}=\mathbf{Af}$ that $\mathbf{f}=\mathbf{Ah}$.

THE REMAINING TRANSFORMATIONS

The transformation relating **f** and **g** can be determined by algebraic manipulation of equation 5 into a form similar to that of equation 6.

$$F(\mathbf{x})=\bigvee_{j=0}^{2^n-1}(\mathbf{f}_j\wedge\mathbf{p}_j) \qquad (5)$$

$$\bar{F}(\mathbf{x})=\bigvee_{j=0}^{2^n-1}(\bar{\mathbf{f}}_j\wedge\mathbf{p}_j)$$

by definition.

$$F(\mathbf{x})=\bigwedge_{j=0}^{2^n-1}(\mathbf{f}_j\vee\bar{\mathbf{p}}_j) \qquad (16)$$

by De Morgan's law extended.

$$F(\mathbf{x})=\bigwedge_{j=0}^{2^n-1}(\mathbf{g}_j\vee\mathbf{q}_j) \qquad (6)$$

But

$$\mathbf{q}_j=\bigvee_{k=0}^{n-1}\left(\mathbf{x}_k\,\Delta\,\mathbf{b}^{j}_{k}\right)=\bigvee_{k=0}^{n-1}\overline{\left(\mathbf{x}_k\,\nabla\,\mathbf{b}^{j}_{k}\right)}$$

and

$$\mathbf{p}_j=\bigwedge_{k=0}^{n-1}\left(\mathbf{x}_k\,\nabla\,\mathbf{b}^{j}_{k}\right)$$

Since $\mathbf{q}_j=\bar{\mathbf{p}}_j$, by De Morgan's law extended, a comparison of equations 16 and 6 yields the result $\mathbf{g}_j=\mathbf{f}_j$, for all j. The transformation relating the canonical forms of equations 5 and 6 is therefore the identity transformation $\mathbf{f}=\mathbf{g}$.

The transformation relating **h** and **i** is found similarly, with the use of the relation

$$\mathop{\nabla}_{j=0}^{r-1}z_j=1\,\Delta\left[\mathop{\Delta}_{j=0}^{r-1}\bar{z}_j\right] \qquad (17)$$

which follows immediately from the definitions of Δ and ∇. Equation 8 can be manipulated algebraically into a form similar to that of equation 7.

$$F(\mathbf{x})=\mathop{\nabla}_{j=0}^{2^n-1}(\mathbf{i}_j\vee\mathbf{s}_j) \qquad (8)$$

$$F(\mathbf{x})=1\,\Delta\left[\mathop{\Delta}_{j=0}^{2^n-1}\overline{(\mathbf{i}_j\vee\mathbf{s}_j)}\right]$$

by equation 17.

$$F(\mathbf{x})=1\,\Delta\left[\mathop{\Delta}_{j=0}^{2^n-1}(\bar{\mathbf{i}}_j\wedge\bar{\mathbf{s}}_j)\right]$$

by De Morgan's law.

$$F(\mathbf{x})=(1\,\Delta\,\bar{\mathbf{i}}_0)\,\Delta\left[\mathop{\Delta}_{j=1}^{2^n-1}(\bar{\mathbf{i}}_j\wedge\bar{\mathbf{s}}_j\right]$$

since $\bar{\mathbf{s}}_0=1$.

$$F(\mathbf{x})=\mathbf{i}_0\,\Delta\left[\mathop{\Delta}_{j=1}^{2^n-1}(\bar{\mathbf{i}}_j\wedge\bar{\mathbf{s}}_j)\right]$$

$$F(\mathbf{x})=\mathop{\Delta}_{j=0}^{2^n-1}\left[\left(\bar{\mathbf{i}}_j\,\Delta\,\mathbf{e}^{0}_{j}\right)\wedge\bar{\mathbf{s}}_j\right] \qquad (18)$$

But

$$F(\mathbf{x})=\mathop{\Delta}_{j=0}^{2^n-1}(\mathbf{h}_j\wedge\mathbf{r}_j) \qquad (7)$$

Since $\bar{\mathbf{s}}_j=\mathbf{r}_j$, by De Morgan's law extended, a comparison of equations 18 and 7 yields the result

$$\mathbf{h}_j=\mathbf{i}_j \, \Delta \, \mathbf{e}^{0}_{j}$$

which may be written more conveniently as

$$\mathbf{h}_j=\mathbf{i}_j \, \nabla \, \mathbf{e}^{0}_{j}$$

The transformations relating the canonical forms of equations 7 and 8 are therefore $\mathbf{h}=\mathbf{i} \nabla \mathbf{e}^0$ and its unique inverse $\mathbf{i}=\mathbf{h} \nabla \mathbf{e}^0$.

The various transformations are summarized below, showing the coefficient vectors of each of the four canonical forms presented, each as a function of each of the others, as follows:

$$\mathbf{f}=\mathbf{g}=\mathbf{Ah}=\mathbf{A}(\mathbf{i} \nabla \mathbf{e}^0)$$

$$\mathbf{h}=\mathbf{Af}=\mathbf{Ag}=\mathbf{i} \nabla \mathbf{e}^0$$

$$\mathbf{i}=\mathbf{h} \nabla \mathbf{e}^0=(\mathbf{Af}) \nabla \mathbf{e}^0=(\mathbf{Ag}) \nabla \mathbf{e}^0$$

References

1. AUTOMATIC DATA PROCESSING (book), K. E. Iverson, F. P. Brooks, Jr. John Wiley & Sons, Inc., New York, N. Y. (to be published).

2. A CHART FOR UNATE FUNCTIONS, W. L. Semon, R. Gould. *Progress Report no. BL-13*, "Theory of Switching," Harvard Computation Laboratory, Cambridge, Mass., Nov. 1955, sect. III, pp. 1–27.

3. ÜBER DIE ERGÄNZUNGSSÄTZE ZU DEN ALLGEMEINEN RECIPROCITÄTSGESETZEN, E. E. Kummer. *Journal fur die reine und angewandte Mathematik*, Berlin, Gormany, vol. 44, 1852, pp. 93–146; also HISTORY OF THE THEORY OF NUMBERS (book), L. E. Dickson. Carnegie Institution of Washington, Washington, D. C., vol. I,1919, pp. 270–71.

4. A CLASS OF MULTIPLE-ERROR-CORRECTING CODES AND THE DECODING SCHEME, I. S. Reed. *Transactions*, Professional Group on Information Theory, Institute of Radio Engineers, New York, N. Y., PGIT-4, Sept. 1954, pp. 38–49.

5. ON A THEORY OF BOOLEAN FUNCTIONS, S. B. Akers, Jr. *Journal*, Society for Industrial and Applied Mathematics, Philadelphia, Pa., vol. 7, Dec. 1959, pp. 487–98.

6. APPLICATION OF BOOLEAN ALGEBRA TO SWITCHING CIRCUIT DESIGN AND TO ERROR DETECTION, D. E. Muller. *Transactions*, Professional Group on Electronic Computers, Institute of Radio Engineers, PGEC-3, Sept. 1954, pp. 6–12.

Sector-Focused Cyclotrons

H. G. BLOSSER
MEMBER AIEE

THIS PAPER will discuss the principles of operation of the sector-focused cyclotron as well as the major design problems which have evolved and the progress which has been made as shown in recent studies of this device. The cyclotron[1] produces energetic beams of charged particles, which are used principally as a research tool to probe the atomic nucleus in an attempt to understand in detail the nature of the forces which hold the nucleus together. An important subsidiary use of the cyclotron beams is for isotope production work. Generally, only neutron-rich radioisotopes can be produced in a nuclear reactor. Neutron-deficient isotopes, of which there are many of practical importance, require the use of charged-particle beams such as are obtained from the cyclotron.

Cyclotron Fundamentals

As the cyclotron is not a conventional piece of electrical engineering equipment, a review of its basic principles of operation will be useful. Consider a charged particle moving at right angles to a uniform magnetic field. The particle moves in a circular orbit with the magnetic force balanced by the centripetal force as indicated in equation 1:

$$F_m=evB=F_{\text{cent}}=mv^2/r$$

$$\omega=v/r=eB/m \tag{1}$$

If this equation is solved for the angular velocity omega, it is seen that the angular velocity is determined entirely by the charge and mass of the particle and the strength of the magnetic field, and is independent of the linear velocity of the particle. In particular, if a pair of radio-frequency (r-f) electrodes is so placed in the region of the particle orbit that the orbit lies in a shielded region during unfavorable parts of the cycle and in a region of strong field during favorable portions of the cycle, and if the radio frequency matches the rotation frequency of the particle, then the particle will remain in phase with the radio frequency and will be accelerated at each passage through the region of strong field. Its linear velocity will therefore steadily increase until it reaches some limiting value which is the maximum the particular instrument is capable of containing. Since the process is fixed frequency, particles of low velocity such as those from a conventional ion source will be accepted on every r-f cycle, and a continuous output beam obtained.

On the basis of the simple picture, the cyclotron is then essentially a perfect accelerator; particles are accelerated continuously (and hence in copious quantities) to arbitrarily high energies. A closer look at details of the process reveals, unfortunately, several severe difficulties which drastically limit the effectiveness of the conventional cyclotron. First, from the theory of relativity, it is known that the mass of a particle is not really constant but depends on the energy of motion in the manner shown in equation 2,

$$m=E/c^2$$

$$\omega=eBc^2/E$$

$$E=(p^2c^2+m_0{}^2c^4)^{1/2}$$

$$p=erB \tag{2}$$

where E is the total energy consisting of the energy of motion plus the rest energy of the particle. Thus, as the particle gains velocity, E increases and hence the mass increases so that there is now a shift in the rotation frequency omega to lower values. This shift can be avoided as shown in equation 3

$$\omega=eBc^2/(e^2r^2B^2c^2+m_0{}^2c^4)^{1/2}$$

$$B(r)=m_0\omega/e\ (1-r^2\omega^2/c^2)^{1/2} \tag{3}$$

in which it is stipulated that the angular velocity shall be a constant, but the magnetic field B is allowed to vary with

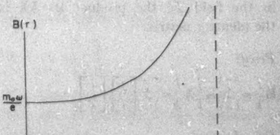


Fig. 1. Isochronous B(r) plotted versus radius

Paper 60-1408, recommended by the AIEE Nucleonics Committee and approved by the AIEE Technical Operations Department for publication. Manuscript submitted October 28, 1959; made available for printing September 29, 1960.

H. G. BLOSSER is with Michigan State University, East Lansing, Mich.

The author wishes to acknowledge his indebtedness to the Oak Ridge National Laboratory and the University of California's Lawrence Radiation Laboratory for permission to use the various figures portraying studies carried out by these two institutions.

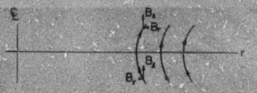


Fig. 2. Form of $B(r, \theta, z)$ when $\partial B_z(r, \theta, z)/\partial r|_{z=0}>0$

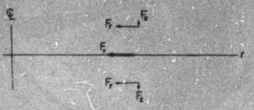


Fig. 3. Force vectors resulting from field of form illustrated in Fig. 2

radius in the cyclotron. With this condition, the first part of equation 3 can be solved for B as a function of r, yielding the result shown. In Fig. 1 the function B of r is plotted versus r, and as the figure shows, is a monatonic increasing function with an infinite singularity at the point where the quantity $r\omega/c$ is equal to 1. This singularity corresponds to rotation of the particle with velocity equal to that of the velocity of light which, on the basis of the theory of relativity, is an unattainable situation. The singularity also corresponds to particles of infinite energy so that for any practical device, the region of interest is the one prior to the singularity, in which the magnetic field rises smoothly from its central value to some maximum specified by the design energy of the particular accelerator. A magnetic field increasing with radius in this fashion (where the rotational frequency is constant) leads, however, to severe difficulties of another character.

To see the nature of these difficulties it is necessary to consider the restrictions which Maxwell's equations impose on the possible shapes of magnetic fields. In the gap of the magnet, the curl of the magnetic field is zero so that a set of equations can be set up relating the various partial derivatives of the magnetic field, as is seen in equation 4.

$$\text{curl } \mathbf{B}=0$$

$$\partial B_r/\partial z=\partial B_z/\partial r \tag{4}$$

If the magnetic field is expanded about the plane $z=0$, equation 5 is obtained. Because of the summetry of the magnetic field with respect to this plane, the z independent term is zero yielding the result

$$B_r(r,\theta,z)=B_r(r,\theta,o)+\partial B_r(r,\theta,z)/\partial z|_{z=0} +\ldots\ldots=z\partial B_z/\partial r \tag{5}$$

If the partial derivative of B_z with respect to r is positive then it is seen at once that the sign of B_r is the same as the sign of z or that the lines of magnetic force have the shape shown on Fig. 2, bending outward away from the center line.

If components of force are computed for particles lying off of the plane of symmetry, or median plane as it is called, the results shown in Fig. 3 are obtained and it is seen that the z component of force is in every case in such a direction as to pull further away from the plane particles which are already out of the plane. Particles exactly on the median plane feel no z force but are in a state of unstable equilibrium. This is a very serious situation. In being accelerated particles must move through a long path from the center of the cyclotron to the extraction region, of the order of miles in a typical cyclotron, and throughout the process must be within 1 inch of the median plane in order to avoid physical obstacles. Since the number of particles lying exactly in the median plane is essentially zero, almost no particles will survive the complete acceleration process if the force situation is as shown in Fig. 3. A field with such properties is referred to by analogy from optics as a defocusing field. Correspondingly, if the off-median plane forces tend to restore the particles to that plane, then the field would be described as focusing. There is, then, a fundamental contradiction between the requirements for focusing and for constant frequency acceleration in the case of magnetic fields with axial symmetry.

In the conventional fixed-frequency cyclotron,[2] this difficulty is resolved in a somewhat unsatisfactory manner by using a magnetic field which decreases with radius to provide the necessary focusing and a high-voltage r-f system which accelerates the particles to their maximum energy before they make a sufficient number of turns to slip out of phase. The maximum energy is then limited by the voltage which can be obtained on the r-f system. The maximum thus far achieved is 400 kv dee to dee in Oak Ridge National Laboratory's (ORNL) 86-inch cyclotron which then yields protons of energy 23 mev (million electron volts).[3]

Sector-Focused Cyclotron Development

PRINCIPLES

In 1938 L. H. Thomas[4] proposed an alternate solution to this dilemma of the fixed-frequency cyclotron. Thomas' solution utilized a magnetic field which varied with azimuth in such a manner that additional focusing forces were derived from the azimuthal variations. These forces could be of sufficient strength to overcompensate the defocusing forces arising from the base magnetic field so that the latter could be set to increase with radius in the manner necessary to provide constant-frequency acceleration. As proposed by Thomas, the variation of the magnetic field with azimuth was of sinusoidal form. It can be shown[5] that the sinusoidal magnetic waveform is not an essential feature of the design; the basic concepts are easier to envisage if one considers a cyclotron utilizing separate uniform-field magnets, as in Fig. 4 which depicts a cyclotron magnet composed of three separate wedge-shaped magnets, each with uniform magnetic field over the region of the wedge and zero field between the wedges. In such a situation, an orbit has the form shown in Fig. 4, being a straight line in the regions between the three wedge magnets and bending through 120 degrees in the regions of magnetic field. A close examination of the region near the edge of a wedge, as shown in section AA of Fig. 4, shows that because of the curvature of the magnetic lines at the edge of the magnet, a particle with trajectory not in the plane of symmetry will encounter an azimuthal component of magnetic field which is opposite in direction, depending upon the side of the median plane in which the particle is moving. Since the particle's trajectory is a straight line in the field-free region between magnets, the velocity v of the particle, if resolved at the point where the particle is entering a magnet, is seen to have both an azimuthal and a radial component. Using the usual cross-product rule, it is found that the cross product of the radial component of velocity and

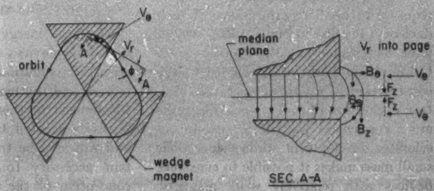


Fig. 4. Origin of sector focusing forces

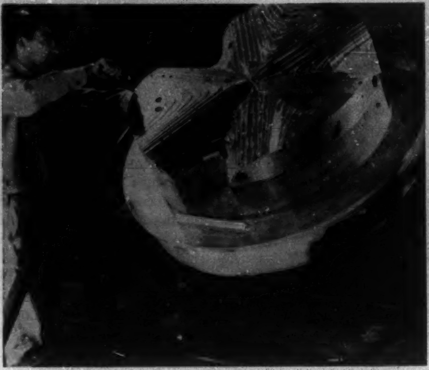


Fig. 5 (left). Assembled pole of University of California Lawrence Radiation Laboratory (UCLRL) 3-sector electron cyclotron

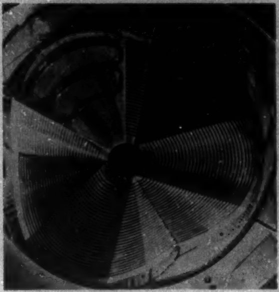


Fig. 6 (right). Orbit coils and three typical "hill" coils used in UCLRL cyclotron

the azimuthal component of the magnetic field is in a direction such as to give a restoring or focusing z force. Likewise, by similar analysis it will be found that a restoring force is also exerted in the region where the particle leaves the magnet; both the radial component of velocity and the azimuthal component of the magnetic field have changed signs, thereby yielding the same sign for the z component of the cross product. These forces at the edge of the magnet then constitute a separate source of restoring forces which can be used to focus the beams of accelerated particles so that they will remain in the desired region and eventually arrive at the extraction radius of the cyclotron. Detailed analysis and numerical calculations show that the strength of this additional focusing force is largely independent of the waveform of the azimuthal variation and, in fact, to good accuracy the focusing strength is found to be specified simply by the mean-square deviation of the field from its mean which can be made as large as necessary.

University of California Model Studies

Although proposed by Thomas in 1938, active development of the sector-focused cyclotron did not begin till much later, principally because of the complexity of the magnetic fields and of the analysis required to predict performance of such devices. Finally, around 1950, studies were undertaken[6] at the University of California in Berkeley which led to the construction of a pair of operating electron cyclotrons[7] of this type. (Electrons are a convenient particle for use in accelerator development studies since their small mass makes it possible to construct devices with relatively weak magnetic fields and small size, but in which the orbit properties are a simple photographic reduction from the orbit properties of devices to accelerate particles of heavier mass.) In these model explorations, three "sectors" were employed in the magnetic field, i.e., the profile of the magnetic field versus azimuth had a threefold symmetry. The azimuthal waveform of the magnetic field was of sinusoidal form in both of the models, the magnet face being intricately machined to a shape which would produce the desired field variation. Fig. 5 shows one of the poles of the magnet. In addition, an extensive arrangement of coils was employed to make empirical fine adjustments to the field, based on observations of beam phenomena. A portion of the coil array is seen in Fig. 6. The models operated successfully, accelerating particles through a large number of turns (of the order of 1,500) which established that uniformity of rotation frequency had been quite accurately achieved. Electrons were accelerated to a final velocity slightly larger than one half of the velocity of light. Fig. 7 shows the beam pattern as seen on a probe being slowly withdrawn from the cyclotron. The extraction of particles at the maximum design energy was also studied in some detail, and considerable success achieved.

Focusing Oscillation Resonances

The final energy of the University of California devices was determined by a resonance condition between the particle motion and the magnet field periodicity. Phenomena of this type have been extensively treated by several authors,[8,11] and will here be treated only to the extent necessary to obtain a qualitative conception of the phenomena involved. If the equations of motion are expanded and solved to yield an equation for the oscillation of a displaced particle about its stable or equilibrium orbit, an equation of the following form is obtained

$$x''+\nu^{2}{}_{x}=\sum a_n\cos n\theta+x\sum b_n\cos n\theta\ldots\ldots \qquad (6)$$

where the a_n, b_n, etc., are related to various harmonic components in the azimuthal variation of the magnetic field, and the quantity ν relates to the frequency of oscillation of the particle about its stable or equilibrium trajectory. From the character of this equation it is seen at once that if ν is equal to any of the integers n, then a resonance condition obtains in which the amplitude of the oscillation will build up in an unbounded manner. Due to the physical limitations on the cyclotron beam space, such a build-up usually leads to loss of the beam. It can be shown that, similarly, terms with coefficient proportional to x cause loss of beam when ν takes on certain half-integral values, and terms with coefficient proportional to x^2 cause loss when ν is equal to certain one-third integral values, etc. This can be understood qualitatively if one considers equation 6, focusing attention on terms of the form x cos$n\theta$, and solves by a successive approximation procedure. In the first approximation the right-hand side of equation 6 is neglected thus yielding the standard solution, $x\sim$ cos$\nu\theta$. Substituting this approximate solution for x in the right-hand side of equa-

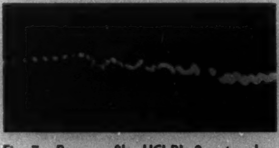


Fig. 7. Beam profile, UCLRL 3-sector electron cyclotron

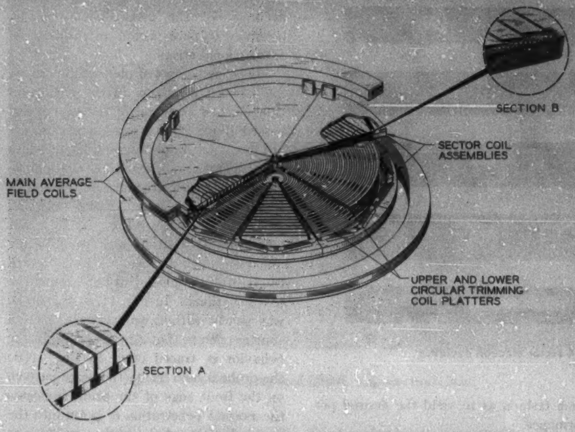


Fig. 8. Schematic drawing of coil structure of ORNL 4-sector electron cyclotron

tion 6 yields terms of the form $\cos\nu\theta\cos n\theta$ which, by standard trigonometric identities, are equivalent to $\cos(n+\nu)\theta+\cos(n-\nu)\theta$. If ν takes on the value $n/2$, then $n-\nu$ is also $n/2$, giving a driving term in resonance with the oscillation and yielding the characteristic rapid buildup of the oscillation amplitude.

For the axial motion of particles in the cyclotron, it is possible by proper manipulation of the amplitude of the sector-focusing fields so to control the oscillation frequency that no resonant value is approached. For the radial motion, this type of control is unfortunately not possible due to the requirement of uniform orbital period for the circulating particles which constrains the radial focusing frequency to a specific pattern of behavior. Thus, in all cases the radial focusing frequency is equal to the orbital frequency at the center of the cyclotron and thereafter rises in approximate direct proportion to the total energy. Hence, if the energy of the cyclotron is sufficiently high, resonant values are eventually encountered.

In the University of California model studies the principal component of the magnetic field was $\cos 3\theta$ so that terms with n equal to 3 were the dominant terms in equation 6. If the radial focusing frequency is measured in units of the orbital frequency, then at the center of the cyclotron it has the value one and thereafter rises, the first integral submultiple of n which is encountered being the value three halves. It was at this point that the acceleration was terminated in the Berkeley models.

More detailed analysis of orbit behavior in the neighborhood of resonances shows that for integral and half-integral resonances an actual stop band exists, by which is meant a band of energies for which there are no stable orbits; 3-sector cyclotrons must then, of necessity, terminate at or before the three-halves point. As the order of the resonance increases, its severity decreases and it is found[8] that for fifth and higher submultiples a region of stability exists even when the frequency of oscillation is exactly equal to the resonant value while for the intermediate third and fourth submultiple resonances, the region of stability vanishes when the frequency takes on exactly the resonant value, but there is no band of unstable frequencies as is the case for the integral and half-integral resonances.

OAK RIDGE NATIONAL LABORATORY MODEL STUDIES

The theoretical analysis leading to the foregoing predictions regarding the behavior of the various types of resonance are of an approximate nature and so in the period 1954–56 studies were undertaken at ORNL to test experimentally several of these predictions. A 4-sector electron cyclotron was constructed[13] which allowed the study of two important resonances. First, it should be noted that the four-halves resonance, which gives the first stop band, now occurred at a radial focusing frequency of 2 so that at intermediate energies the beam would pass through the next higher order essential resonance which occurred at a frequency

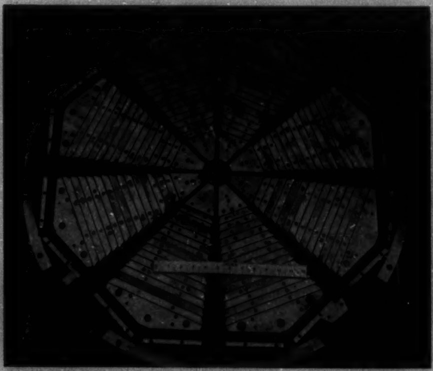


Fig. 9. Set of azimuthal coils for ORNL 4-sector electron cyclotron

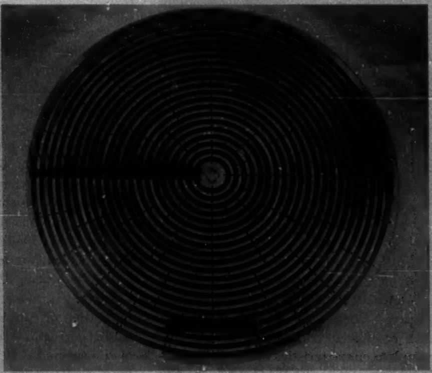


Fig. 10. Set of circular coils for ORNL 4-sector electron cyclotron

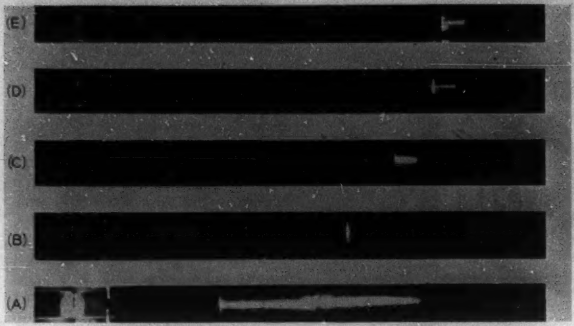


Fig. 11. Beam patterns for ORNL 4-sector electron cyclotron

of 4/3. In addition, although the magnetic field was designed to have purely fourth and higher harmonics, it was of interest to test the effect of the small amounts of third harmonic which would occur in the magnetic field as imperfection components. These imperfection components would manifest themselves as a three-halves resonance, which is therefore termed an "imperfection resonance."

The Oak Ridge electron cyclotron was of air-cored construction, i.e., no ferromagnetic materials were employed. Fig. 8 is a schematic diagram of the coil structure used in the device. Sets of circular coils provided a base field which increased with radius in the manner necessary to compensate the increase in mass of the particles as they were accelerated. Superimposed on this base field were azimuthal variations formed from the wedge-shaped sector windings seen in the figure. Fig. 9 is an actual photograph of a set of the wedge-shaped windings assembled and ready to be installed in the cyclotron. It will be noticed that there are eight coils, one coil of each pair of wedge sections being used to produce a hill or strong region of field, the adjacent coil being used to produce a valley or weak region of the field. Fig. 10 shows the layer of circular trimming coils which were used in conjunction with a large main field coil to produce the proper increase of field with radius. In the design process, the magnetic field produced by each of the coils was numerically computed and utilized for prior study of the orbit properties by means of numerical integration of the equations of motion. With a digital computer it was then possible to compute in advance the complete characteristics of the beam performance in the cyclotron and to modify the design in advance of construction in such fashion as to yield the desired performance.

Utilization of the digital computer for studies of this type has, in fact, become characteristic of recent development studies on the sector-focused cyclotrons, and is probably more responsible than any other factor for recent rapid progress in the development of such machines. Computer routines capable of tracking the trajectory of a particle through complex magnetic and electric field configurations have been developed by a number of persons,[13,14] and for several different types of computers. Such routines make it possible to study the behavior of orbits in close detail; with such information, the type of alteration necessary to effect improvements in performance is usually self-evident or becomes so after study of a few trial alterations.

The ORNL electron cyclotron was the first sector-focused accelerator to be completely designed on the basis of precomputed magnetic fields and orbit properties. Construction of the device consisted of fabrication of coils in careful conformity with computed specifications. Construction was completed and the cyclotron put into operation without recourse to mapping of the magnetic field, which is always a laborious process.

As a check on the reliability of such procedures, the performance of the device was experimentally studied in some detail after construction. Electrons in the machine made as many as 2,000 revolutions before reaching the extraction radius, indicating that a precision of the order of 1 in 5,000 had been achieved in realization of constant orbital frequency. Radial and axial focusing frequencies were also measured. In each case the measured performance duplicated the computed predictions to within either the accuracy of the measurement or the accuracy to be expected on the basis of fabricational errors.

Fig. 11 is a series of photographs showing beam phenomena in the cyclotron; they illustrate several of the effects discussed above. Each band across the figure is a separate photograph looking into the cyclotron. In each of the photographs, particles originate at the source which is on the far left of the band and are accelerated until they strike a flat metal probe coated with scintillating material which is seen to shine toward the right of the photograph in each of the five situations. In Fig. 11(A) the probe was slowly withdrawn with the camera shutter open so that a radial profile of the behavior is traced out. In Fig. 11(B) the probe is held fixed; the beam pattern on the front edge of the probe indicates the normal penetration of beam into the probe due to radial focusing oscillations and to acceleration. In Fig. 11(C) the penetration of the beam into the probe has greatly increased due to effects of the four thirds resonance which has excited large focusing oscillations in part of the beam, causing this portion to penetrate deeply into the probe. In Fig. 11(D) an additional view of the beam somewhat further out is shown in which the separation of the beam into two components is quite clearly evidenced. The "normal" or undisturbed component of the beam gives a bright vertical line on the front edge of the probe similar to that in Fig. 11(B), while the horizontal line extending across the probe is produced by particles of very-large-amplitude radial oscillation, the oscillation having been excited in passing through the resonance. In Fig. 11(E) a careful look will show that the "normal" component of the beam has now also begun to stream across the probe which indicates that the particles have reached the maximum energy and are moving into the band of instability at radial focusing frequency 2. Effects of the three-halves imperfection resonance are too small to be discerned. The studies of resonant phenomena which were made on this electron model showed agreement with the theoretically predicted behavior; this was a very important result and indicated at once that one of the most serious problems in cyclotrons of this type was of a tractable character.

HIGH-ENERGY SECTOR-FOCUSED-CYCLOTRON PROPOSAL

As a result of this work, the Oak Ridge group has made a study of a very large cyclotron of the sector-focused type utilizing an 8-sector magnetic field to accelerate

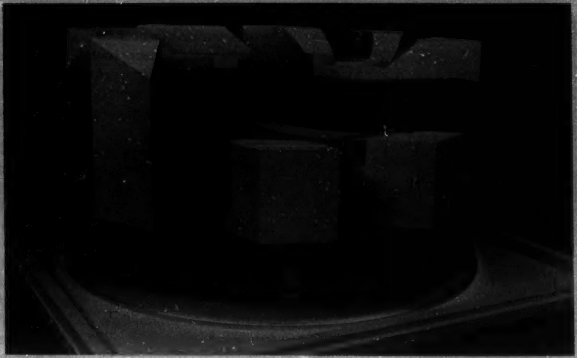


Fig. 12. Model of proposed ORNL 850-mev cyclotron

protons to a final energy of 850 mev.[15] Fig. 12 shows a model of this proposed cyclotron. The over-all diameter is approximately 50 feet, and about 2,500 tons of iron are required for construction of the magnet. Acceleration would be provided by two S-shaped r-f cavities which thread through the machine, the cavities being excited in the transverse electric mode, with the particles passing in and out of the cavity through thin slots cut along nodes of the cavity current distribution. Both the r-f cavities and the magnet would be enclosed in a large cylindrical chamber to obtain the necessary vacuum. It will be noted in the illustration that the magnet sectors are spiraled, an important effect not mentioned up to this point.

The influence of spiral sectors in strengthening the net axial focusing was first pointed out by D. W. Kerst of the Midwestern Universities Research Association.[16] Introduction of spiral clearly greatly strengthens the focusing at one edge of a sector and correspondingly weakens the force at the opposite edge with the result that one edge is strongly focusing and the other rather strongly defocusing. This is just the situation which is necessary to obtain the so-called "strong" or "alternating gradient" focusing which is being employed in several large synchrotrons now in construction. Details of the effect have been extensively discussed elsewhere[17] and so will not be considered further here. (For the reader who desires a qualitative understanding of the effect a review of the corresponding optical situation is suggested. It will be found that a positive and negative lens of equal focal length when separated by some distance L give rise to a net positive focusing force of considerable strength because a given ray passes through the positive lens further from the axis than when passing through the negative lens and so receives a larger positive bend.)

The feasibility of this very large spiral-ridge cyclotron is presently being further explored at Oak Ridge. A second electron model is in design which will realistically incorporate essential features of the large 8-sector machine.

MEDIUM-ENERGY SECTOR-FOCUSED CYCLOTRONS

Subsequent to the University of California and Oak Ridge model studies of high-energy sector-focused cyclotrons, considerable interest has arisen in the application of the principle in the medium-energy region (the region extending from approximately 30 to 100 mev). This region is just above the boundary of the region accessible to conventional fixed-frequency cyclotrons (as mentioned above, the highest particle velocity which has been achieved in a cyclotron of this type is in the Oak Ridge 86-inch cyclotron which yields 23-mev protons). Coupled with the use of sector focusing to extend the energy range has been a movement toward greater flexibility and variability of the beam characteristics of the cyclotron, which results in research tools of unusually versatile character.

The first major step along these lines was made at the Los Alamos Scientific Laboratory in a major redesign and rebuilding of the Los Alamos 42-inch cyclotron[18] which was completed about 2 years ago. Small sector-focusing shims were incorporated in the cyclotron, adding somewhat to the focusing strength which was then derived from a combination of conventional cyclotron focusing plus sector focusing. The base field in the cyclotron still decreases with radius so only a limited number of turns can be accomplished before the particles slip out of phase, the situation being analogous to that which would obtain in a strictly conventional cyclotron. In addition to the sector shims, the cyclotron was provided with an oscillator of adjustable frequency and with devices to permit variation of the magnetic field over a considerable range of field strength, so that it became possible to vary the energies of the output particles and also to accelerate different particles. Prior to the Los Alamos work a small conventional cyclotron with adjustable magnetic field and radio frequency had been constructed at the University of Rochester and a second larger instrument at the University of California's Livermore Laboratory.[19]

Recently a group of Russian scientists reported the completion of a sector-focused spiral-ridged cyclotron at the government laboratory in Dubna.[20] In this cyclotron, the main field increases with radius as is required to provide for constant-frequency acceleration and the machine thus is one of the first operating cyclotrons depending entirely on sector focusing to contain the particles.

Similar devices are in design or construction at a number of laboratories in the United States. In particular, the

Fig. 13. Magnet yoke for 75-mev Oak Ridge relativistic isochronus cyclotron

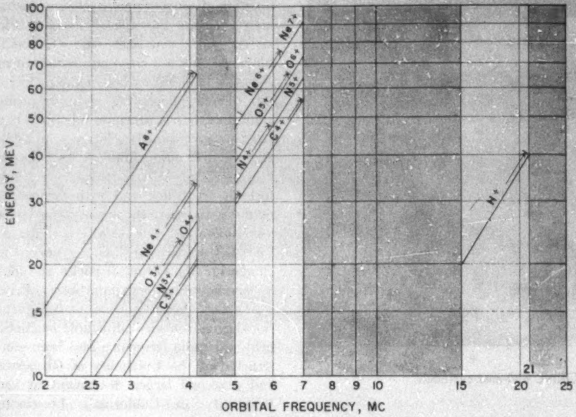


Fig. 14. Particle energy versus orbital frequency for the various particles and energies to be obtained from proposed MSU 64-inch cyclotron

laboratories engaged in actual construction of new cyclotrons of this type are the Oak Ridge National Laboratory, The University of California's Lawrence Radiation Laboratory, University of California at Los Angeles, and the University of Colorado. It seems probable that one or more of these machines will be in operation late in the year 1961. Fig. 13 shows the magnet which will be used in the Oak Ridge cyclotron and gives some idea of the size of equipment involved.

DESIGN FEATURES OF PROPOSED MICHIGAN STATE UNIVERSITY CYCLOTRON

In addition to the above listed machines which are in construction, a number of groups are engaged in the active design of machines of this type, including a group at Michigan State University (MSU) of which the author is a member. In order to provide some idea of the properties of machines of this type, the major features of the proposed MSU cyclotron will be reviewed. This machine is, in most respects, similar to machines being designed or constructed at other institutions and so will be a reasonably typical example of the type of instrument. It also, of course, is the one with which the author is most familiar.

Fig. 14 is a plot of the particles and energies which it is expected to obtain from the MSU cyclotron. The r-f system will operate over the range from 15 to 21 mc which will, in turn, by virtue of the properties of the cyclotron, accelerate particles on any odd submultiple of the radio frequency so that particles or orbital frequency from 5 to 7 mc are accelerated on the third submultiple and similarily for lower orbital frequencies. On the fundamental frequency, protons of energy adjustable in the range 20 to 40 mev are obtained. On the lower frequency a variety of so-called "heavy ions" are accelerated to a broad range of energies; where the term "heavy ions" refers to partially ionized nucleii of atoms in the carbon, nitrogen, oxygen, neon range of the periodic table.

Fig. 15 is a diagram of the pole-tip profile which will be used in the magnet to provide the sector-focusing forces. Of particular interest is the relatively simple geometric configuration of the pole tip; the simple geometry results from taking full advantage of the arbitrary character of the azimuthal waveform of the magnetic field. This pole is in marked contrast to the complex pole structure required to achieve an explicit sinusoidal dependence for the azimuthal variation. The simplification of the pole structure from that thought necessary in the early work is one of the major factors in the evolution of the sector-focused cyclotron to a state of proved feasibility.

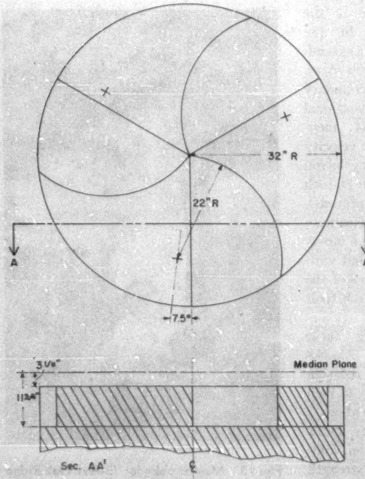


Fig. 15 (left). Pole-tip geometry of proposed MSU 64-inch cyclotron

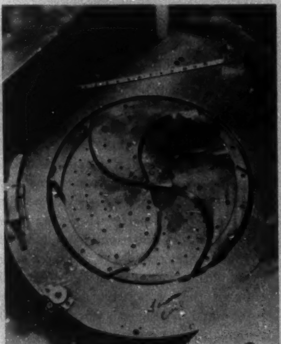


Fig. 16 (right). Model magnet used in design studies of UCLRL 88-inch cyclotron

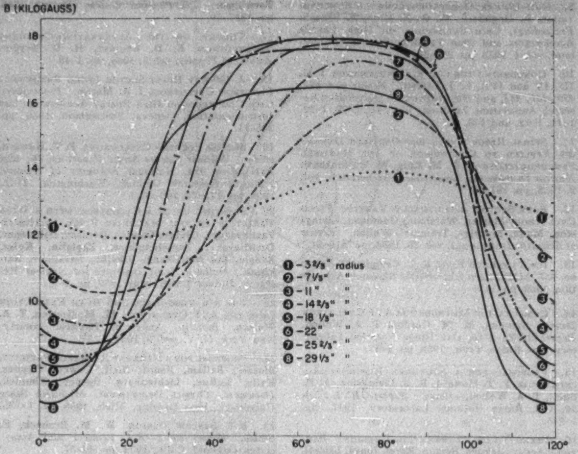


Fig. 17. Magnetic field obtained from pole tip of the type shown in Fig. 15

Fig. 16 shows one of the pole geometries being considered for use in the UCLRL 88-inch cyclotron now in construction. The illustration shows the 1/5-scale model which is being used in the studies. It will be noted that considerably greater spiral is employed than in the geometry shown in Fig. 15.

The actual variation of magnetic field with azimuth produced by the pole-tip shape of Fig. 15 is shown in Fig. 17. It is seen that the field varies from roughly 8 to 18 kilogauss in going from valley to hill at radii near the maximum. This is sufficient to provide adequate focusing for all the proposed particles. Fig. 18 shows the computed axial focusing frequency for protons and a typical heavy ion. Such studies of the properties of cyclotrons of this type are carried out by constructing accurately scaled models of the proposed magnet and measuring the field produced by the model. The measured field information is used to make digital computations of the properties of orbits. In this way it can be established whether the focusing is adequate and the rotational frequency sufficiently constant to provide the desired acceleration of particles to the maximum energy.

Considerable study has been made of the possibility of using one of the resonances in such a device as a means for extraction of the beam of particles. Conceptually, such an extraction is accomplished by so perturbing the field that a resonance is reached at the maximum energy. Particles reaching this energy then undergo the typical resonant increase in the amplitude of their focusing oscillation so that a very large gain in amplitude occurs, the gain being large enough in some situations to carry the particles completely out of the machine, thus accomplishing the extraction with a minimum of additional physical components. Extensive studies of processes of this sort have been carried out at Michigan State and other institutions and further studies are in progress.[21,22] At the present state of progress, it is well established that systems of this type are feasible, although complete details have not yet been worked out.

Fig. 19 shows in schematic form the r-f system[23] for the MSU cyclotron. The D-shaped accelerating electrodes are seen at the lower part of the figure. These electrodes constitute a capacitive load which is coupled through an inductive line to yield a system resonant at the required frequency. Adjustment of the resonant frequency is accomplished by variation of the characteristic impedance of the inductive line, the variation of impedance being achieved by an actual mechanical shift in the shape of the external conductor of the line using a system of flexible panels. In this way a system resonant over the range from 15 to 21 mc can be achieved with high Q such that the dee's can be excited to 140 kv by an r-f power source of approximately 200 kw. Other features are of less importance and will not be discussed further in this paper.

Conclusions

It is expected that in the United States during the next 5 to 10 years approximately a dozen cyclotrons will be built of the medium-energy sector-focused type such as the one described above. In addition, there is a distinct possibility that one or more of the high-energy cyclotrons discussed earlier in the paper will also be built. These devices will constitute powerful new tools for use in

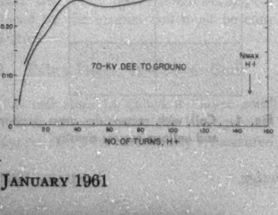
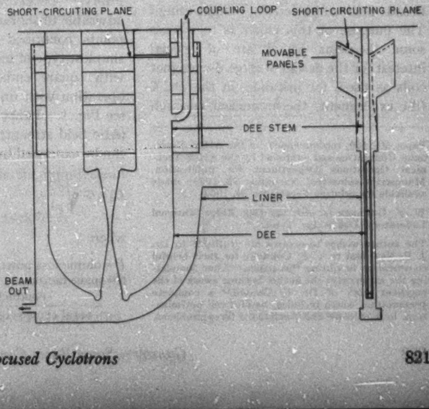


Fig. 18 (left). Calculated axial focusing frequencies of two typical particles in a magnetic field of the form shown in Fig. 17

Fig. 19 (right). Sketch of dee and liner assembly of proposed MSU 64-inch cyclotron



furthering our knowledge of the atomic nucleus and nuclear phenomena.

References

1. On the Production of High Speed Protons, **E. O. Lawrence, N. E. Edelfsen.** *Science*, New York, N. Y., vol. 72, 1930, pp. 376–77.

2. Cyclotrons and Synchrocyclotrons (book), **B. L. Cohen.** Springer-Verlag, Berlin, Germany, vol. 44, 1959, pp. 105–53.

3. The Oak Ridge 86-Inch Cyclotron, **R. S. Livingston.** *Nature*, London, England, vol. 170, 1952, pp. 221–32.

4. The Paths of Ions in the Cyclotron, **L. H. Thomas.** *The Physical Review*, New York, N. Y., vol. 54, 1938, pp. 280–98.

5. A Fixed Frequency Billion Volt Proton Cyclotron, **T. A. Welton.** *Ibid.*, vol. 99, 1955, p. 1623.

6. Theoretical Study of Relativistic Constant Frequency Cyclotron, **D. L. Judd.** *Ibid.*, vol. 100, 1955, p. 1804.

7. Two Electron Models of a Constant–Frequency Relativistic Cyclotron, **Kelly, Pyle, Thornton, Richardson, Wright.** *Review of Scientific Instruments*, New York, N. Y., vol. 27, 1956, pp. 492–503.

8. Non–Linear Effects in Synchrotrons, **P. A. Sturrock.** *Annals of Physics*, New York, N. Y., vol. 3, 1958, pp. 113–89.

9. Non–Linear Orbit Problems in Synchrotrons, **R. Hagedorn, M. G. N. Hine, A. Schoch.** *Proceedings*, Cern Symposium on High Energy Accelerators and Pion Physics, Geneva, Switzerland, vol. 1, 1956, pp. 237–53.

10. Concerning the $\nu/n \rightarrow 1/3$ Resonance, I, II, III, IV and IVA, **L. J. Laslett.** *Report nos. 452, 459, 461, 463,* and *490*, Midwestern Universities Research Association, Madison, Wis., pp. 1–39, 1–22, 1–14, 1–12, and 1–6.

11. Spiral Ridge Cyclotron Particle Dynamics Applied to Conversion of the Harwell Synchrocyclotron, **N. M. King, W. Walkinshaw.** *Nuclear Instruments*, Amsterdam, Netherlands, vol. 2, 1958, pp. 287–98.

12. Four-Sector Azimuthally Varying Field Cyclotron, **Blosser, Worsham, Goodman, Livingston, Mann, Moseley, Trammel, Welton.** *Review of Scientific Instruments*, vol. 29, 1958, pp. 819–34.

13. Ill Tempered Five, **E. Z. Chapman.** *Report no. 457*, Midwestern Universities Research Association, 1959, pp. 1–34.

14. Computation Methods for AVF Cyclotron Design Studies, **M. M. Gordon, T. A. Welton.** *Report ORNL-2756*, Oak Ridge National Laboratory, Oak Ridge, Tenn., 1959, pp. 1–47.

15. Proposal for a Southern Regional Accelerator, **F. T. Howard, R. S. Livingston, A. H. Snell, T. A. Welton,** editors. *Report ORNL 57-4-30*, Oak Ridge National Laboratory, 1957, pp 1–245.

16. Fixed–Field Alternating-Gradient Particle Accelerators, **Symon, Kerst, Jones, Laslett, Terwilliger.** *The Physical Review*, vol. 103, 1956, pp. 1837–59.

17. Theory of the Alternating–Gradient Synchrotron, **E. D. Courant, H. S. Snyder.** *Annals of Physics*, vol. 3, 1958, pp. 1–48.

18. A 450-kev Eight-Sector Fixed Frequency Electron Cyclotron, **J. A. Martin.** *Proceedings*, Cern Conference on High Energy Accelerators and Instrumentation, Geneva, Switzerland, 1959, pp. 205–11.

19. Sector Focused Cyclotrons, **F. T. Howard,** editor. *Nuclear Science Series Report no. 26*; also *Publication 656*, National Academy of Science, National Research Council, Washington, D.C., 1959, pp. 171–74, 262–66.

20. Starting Up of a Cyclotron with a Wide Variation in the Magnetic Field Intensity, **Vasilevskaya, Glazov, Danilov, Denisov, Dzhelepov, Dmitrievsky, Zamolodnikow, Zaplatin, Kolga, Kropin, Liu Nai-Chang, Ribalko, Savenkov, Sarkisian.** *Bulletin*, Joint Institute for Nuclear Research, Dubna, USSR, May 1959, pp. 1–8.

21. The 8/4 Resonance and Beam Extraction from the AVF Cyclotron, **M. M. Gordon, T. A. Welson.** *Bulletin*, American Physical Society, New York, N. Y., vol. 3, 1959, p. 57.

22. Proposal for a Nuclear Research Facility, **Blosser, Ballam, Beard, Blatt, Cowen, Haynes, Kelly, LaRue, Lichtenberg, Spence, Timnick,** *Document*, Physics Department, Michigan State University, East Lansing, Mich., 1958, pp. 1–135.

23. R-F System Design, **W. M. Brobeck, B. Smith.** *Report 200-18-R-1*, Brobeck and Associates, Oakland, Calif., 1958, pp. 42–57.

Some Basic Concepts for Magnet Coil Design

W. F. GAUSTER
MEMBER AIEE

IN THE United States and abroad there is intensive activity in thermonuclear research.[1] A major part of this effort is directed toward containment, in strong magnetic fields, of plasma of the necessary high-particle energies, corresponding to temperatures of hundreds of millions of degrees Kelvin. Magnetic field design in thermonuclear research has already been developed to a relatively broad field of electrical engineering.[2] The purpose of this paper is to discuss some theorems which are of special interest for the design of large d-c magnet coils as used, for instance, in the DCX (d-c experiment) thermonuclear research program of the Oak Ridge National Laboratory.[3]

1. Cylindrical Magnet Coils With Square Ends and Uniform Current Density

The design of high-power magnet coils is based on fundamental work by Fabry,[4,5] Cockcroft,[6] and Bitter.[7] It is advisable to summarize a few well-known points contained in this theory. The simplest type of magnet coil is cylindrical with square ends and is designed for operation with uniform current density; see Fig. 1. Fabry showed that the magnetic field strength at the coil center H_0 can be expressed by

$$H_0 = G\sqrt{\frac{P\lambda}{\rho a_1}} \tag{1}$$

where

P = ohmic loss power, watts

λ = space factor $= \dfrac{\text{conducting cross section}}{\text{total cross section including insulation}}$

ρ = resistivity of the coil conductor, ohm-cm (ohm-centimeters)

a_1 = inside radius of the coil, cm

H_0 is measured in oersteds, and B_0 has the same numerical value in gausses (no ferromagnetic material is supposed to be in the magnetic field). G is a dimensionless factor, the so-called Fabry factor, which does not depend on the size but only on the shape of the coil.

By introducing the ratios

$$\alpha = \frac{a_2}{a_1};\quad \beta = \frac{b}{a_1} \tag{2}$$

(see Fig. 1), G can be expressed as a function of α and β in an elementary way.[4] It is convenient to represent G by a family of curves; see Fig. 2.[8] A "flat"

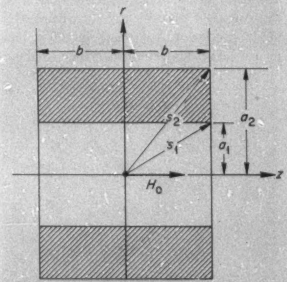


Fig. 1. Coil with rectangular cross section and uniform current density

---

Paper 60-1409, recommended by the AIEE Nucleonics Committee and approved by the AIEE Technical Operations Department for publication. Manuscript submitted November 25, 1959; made available for printing September 29, 1960.

W. F. Gauster is with the Oak Ridge National Laboratory, Oak Ridge, Tenn.

The author wishes to express his gratitude to Dr. J. P. Neal and to T. F. Connolly for their helpful co-operation in editing this paper. After completing the manuscript the author became aware of the excellent work of Dr. F. Gaume,[13] a complete presentation which includes, besides coil optimization, heat transfer and mechanical force problems.

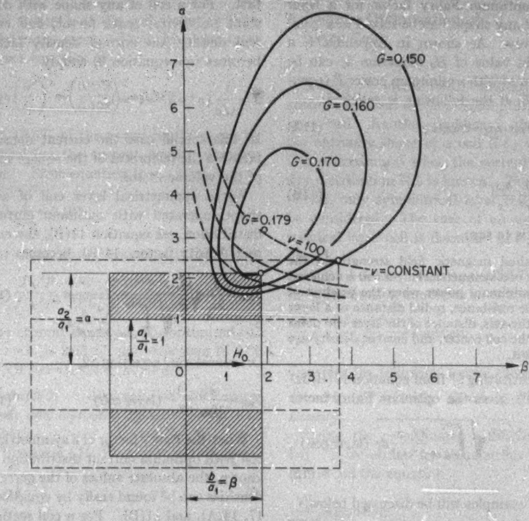


Fig. 2. Fabry and volume factor diagrams of cylindrical magnet coils with square ends and uniform current density

maximum of $G = 0.179$ occurs for $\alpha = 3.09$, $\beta = 1.88$.

Fabry also used a volume factor v, defined by

$$v=\frac{\text{total coil volume}}{a_1^3} \tag{3}$$

For instance, in the case of a cylindrical coil with square ends, $v = 2\pi(\alpha^2-1)\beta$. In Fig. 2, v = constant curves are shown by dashed lines. $G_{\max}$ corresponds to $v=101$. A somewhat smaller G can be achieved with a much smaller volume factor. For instance, the DCX coils are designed with the following data:[9]

$\alpha=1.53$
$\beta=0.618$
$v=5.21$
$G=0.128$

Fabry gave a table showing, for volume factors between 1 and 10^5, values of α and β for which G assumes maximal values.[10] The DCX coil is such a minimum-volume coil. It should be pointed out, however, that for many applications a minimum volume factor is not of special importance, since the whole field configuration produced by the magnet coil must be considered.

2. Fabry Factor of a Layer Coil

A coil shall be called a "layer coil" if it is wound in straight homogeneous layers; i.e., if current density i, measured in amp/cm² (amperes per cm²), resistivity ρ, and space factor λ are functions only of the distance from the axis of r. If these quantities are functions of r and z, the coil will be called a "loop coil." Fig. 3 shows a layer coil which is symmetrical to a plane through 0 perpendicular to the z-axis. In the following, a general expression for the Fabry factor G will be derived for a symmetric layer coil where the magnetic field strength H_0 is referred to the coil center. A similar expression can be found for an asymmetric layer coil, and H_0 referred to any point at the axis; see Appendix II.

A cylindrical current sheet with a radius r, the length $2z$, and a thickness dr produces in the center the field strength (compare, for instance, reference 11).

$$dH_0=\frac{4\pi}{10}\lambda i\cos\phi\, dr=\frac{4\pi}{10}\lambda i\frac{z}{s}\,dr$$

Therefore

$$H_0=\frac{4\pi}{10}\int_{r=a_1}^{a_2}\lambda i\frac{z}{s}\,dr \tag{4}$$

The power dissipated in the infinitesimal current sheet is

$$dP=\rho i^2 dV_{Cu}=\rho\lambda i^2 dV=4\pi\rho\lambda i^2 rz\,dr$$

Therefore

$$P=4\pi\int_{r=a_1}^{a_2}\rho\lambda i^2 rz\,dr \tag{5}$$

Normalized (dimensionless) quantities, which are in general functions of r, are introduced as follows:

$$i^+=\frac{i}{i_1};\ \rho^+=\frac{\rho}{\rho_1};\ \lambda^+=\frac{\lambda}{\lambda_1} \tag{6}$$

The subscript 1 refers to the innermost layer where $r=a_1$.

Equations 4 and 5 become

$$H_0=\frac{4\pi}{10}\lambda_1 i_1\int_{r=a_1}^{a_2}\lambda^+ i^+\frac{z}{s}\,dr \tag{4A}$$

and

$$P=4\pi\rho_1\lambda_1 i_1^2\int_{r=a_1}^{a_2}\rho^+\lambda^+(i^+)^2 rz\,dr \tag{5A}$$

Eliminating i_1 from these equations (compare Bitter's computation of G for several special cases[7]) yields

$$H_0=G\sqrt{\frac{P\lambda_1}{\rho_1 a_1}} \tag{7A}$$

$$G=\sqrt{\frac{\pi a_1}{5}}\,\frac{\int_{r=a_1}^{a_2}\lambda^+ i^+\frac{z}{s}\,dr}{\sqrt{\int_{r=a_1}^{a_2}\rho^+\lambda^+(i^+)^2 rz\,dr}}$$

(layer coil) **(7B)**

Equation 7(B) contains a_1; however, G is dimensionless and therefore independent of the magnitude of a_1.

3. Fabry Factors of Cylindrical Coils with Tapered Ends

As an example of the application of equations 7(A) and 7(B), a cylindrical coil with tapered ends (see Fig. 4), constant space factor ($\lambda^+=1$), and constant specific resistance ($\rho^+=1$) will be considered. The current density is also assumed to be constant, inversely proportional to the radius r, or inversely proportional to r^2.

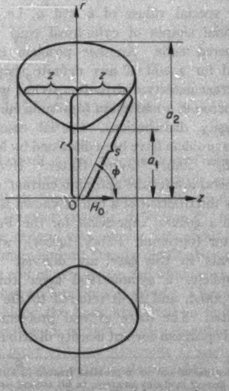


Fig. 3. Symmetrical layer coil

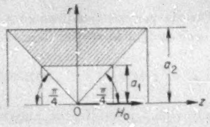


Fig. 4. Cylindrical coil with tapered ends

Case 1: $i^+=1$

$$G=\frac{\sqrt{3\pi}}{5}\sqrt{\frac{k}{1+k^2}}\frac{\alpha-1}{\sqrt{\alpha^2-1}} \tag{8}$$

The maximum value for G is 0.172, for an angle $\phi=45$ degrees ($k=1$), and $\alpha=2.7$. This is case 4 of Bitter's paper.[7]

Case 2: $i^+=a_1/r$

$$G=\frac{\sqrt{\pi}}{5}\sqrt{\frac{k}{1+k^2}}\frac{\ln\alpha}{\sqrt{\alpha-1}} \tag{9}$$

The maximum value for G is 0.201, again for an angle $\phi=45$ degrees, but for $\alpha=4.5$. This is Bitter's case 5.

Case 3: $i^+=(a_1/r)^2$

$$G=\frac{\sqrt{\pi}}{5}\sqrt{\frac{k}{1+k^2}}\sqrt{\frac{\alpha-1}{\alpha}} \tag{10}$$

The maximum value for G is 0.250, for an angle $\phi=45$ degrees, and α approaching infinity. This case has been previously discussed by Gaume.[12] It deserves interest because this current distribution yields the maximum G for a cylindrical layer coil with tapered ends.

4. Optimum Current Density Distribution of a Layer Coil

In the previous paragraph maximum values of G were considered which occur for special values of k and α, i.e., for special shapes of cylindrical coils with tapered ends. Another problem is to find for a coil of any definite shape a current density distribution $i(r)$ at which G becomes a maximum (optimum current density distribution). Special cases of this problem have been discussed by Maxwell,[13] Fabry,[*] and Bitter.[7] Here a general law for the optimum current density distribution of a coil of known shape and a general expression for the Fabry factor (optimum Fabry factor), which results in this case, are derived. For simplicity a symmetrical layer coil is supposed, and H_0 is referred to the coil center. The more general problem, to find optimum current density distribution and optimum Fabry factor for a layer coil of any shape, can be solved in a similar way. As shown in Appendix I, a certain value of H_0 (equation 4) can be produced with a minimum power P (equation 5) if the following is satisfied:

$$\rho r s i=\rho_1 a_1 s_1 i_1=\text{constant} \tag{11A}$$

Thus

$$i^+=\frac{1}{\rho^+}\frac{a_1 s_1}{s} \tag{11B}$$

That is to say:

A desired magnetic field strength in the center of a symmetrical layer coil is achieved with minimum power when the product of specific resistance, radial distance of a layer from the axis, distance of the layer end point from the coil center, and current density are constant.

Substituting i^+ from equation 11(B) into 7(B), gives the optimum Fabry factor

$$G_{\text{opt}}=\frac{\sqrt{\pi a_1}}{5}\sqrt{\int_{r=a_1}^{a_2}\frac{\lambda^+}{\rho^+}\frac{z}{rs^2}\,dr}\ \text{(layer coil)} \tag{12}$$

Some examples will be discussed below.

5. Current Density Factor J of a Layer Coil

As shown, the Fabry factor G is dimensionless and characteristic of a certain coil shape and the distribution of relative current density, space factor, and resistivity. The absolute value of H_0 is found by multiplying G by a scaling factor:

$$\sqrt{\frac{P\lambda_1}{\rho_1 a_1}}$$

See equation 7(A).

A similar procedure may now be used to find the absolute value of the current density i_1. From equation 5(A)

$$i_1=\sqrt{\frac{P}{4\pi\rho_1\lambda_1\int_{r=a_1}^{a_2}\rho^+\lambda^+(i^+)^2 rz\,dr}} \tag{13}$$

This equation can be written as

$$i_1=J\sqrt{\frac{P}{\rho_1\lambda_1 a_1^3}} \tag{14A}$$

$$J=\frac{1}{2}\sqrt{\frac{a_1^3}{\pi}}\sqrt{\frac{1}{\int_{r=a_1}^{a_2}\rho^+\lambda^+(i^+)^2 rz\,dr}} \tag{14B}$$

The dimensionless factor J shall be called the "current density factor." The absolute value of i_1 is found by multiplying J by a scaling factor:

$$\sqrt{\frac{P}{\rho_1\lambda_1 a_1^3}}$$

The analogy to the Fabry factor is obvious.

Two special cases might be considered first. For a coil of any shape with constant resistivity, space factor, and current density, the current density factor becomes (see equation 3) simply

$$J=\frac{1}{\sqrt{v}}\ (\rho^+=\lambda^+=i^+=1) \tag{15}$$

In this special case the current density factor is the reciprocal of the square root of the volume factor.

For a symmetrical layer coil of any shape, operated with optimum current distribution, see equation 11(B), the current density factor, 14(B), becomes

$$J_{\text{opt}}=\frac{1}{2s_1}\sqrt{\frac{a_1}{\pi}}\frac{1}{\sqrt{\int_{r=a_1}^{a_2}\frac{\lambda^+}{\rho^+}\frac{z}{rs^2}\,dr}} \tag{16}$$

Or, considering equation 12,

$$J_{\text{opt}}=\frac{a_1}{10s_1}\frac{1}{G_{\text{opt}}}\ \text{(layer coil)} \tag{17}$$

When the Fabry factor of a symmetrical coil with optimum current distribution is known, the absolute values of the current densities can be found easily by equations 17, 14(A), and 11(B). For a coil section similar to that shown in Fig. 3, $a_1=s_1$, and therefore

$$J_{\text{opt}}=\frac{1}{10}\frac{1}{G_{\text{opt}}}\ \text{(layer coil)},\ a_1=s_1 \tag{17A}$$

6. Examples of Layer Coils with Optimum Current Density Distribution

Example 1. Cylindrical Coil with Square Ends and Optimum Radial Current Distribution, $\rho^+=\lambda^+=1$; *see Fig 1*[5,7]

In this case equation 11(B) becomes

$$i^+=\frac{a_1\sqrt{a_1^2+b^2}}{r\sqrt{b^2+r^2}} \tag{18}$$

and equation 12 yields

$$G_{\text{opt}}=\frac{1}{10}\sqrt{\frac{2\pi}{\beta}\ln\frac{\alpha^2(1+\beta^2)}{\alpha^2+\beta^2}} \tag{19}$$

For a coil with $\alpha=3$, $\beta=2$, the Fabry factor is $G_{\text{opt}}=0.197$. From equation 17 the current density factor is

$$J_{\text{opt}}=\frac{1}{10\sqrt{5}}\times\frac{1}{0.197}=0.226$$

Equation 11(B) yields for the current density at $r=a_2=3a_1$

$$i_2=\frac{a_1 s_1 i_1}{a_2 s_2}=\frac{\sqrt{5}}{3\sqrt{13}}i_1=0.206i_1$$

It is interesting to compare these data with those for a cylindrical coil with square ends and constant current density and with the same a_1, α, β, ρ, and λ. The

---

[*] One special optimum problem treated in Fabry's first paper[4] contains an error; in his second paper,[5] however, all optimum values indicated are correct.

Vol
Pt.
196

field strength at the coil center in both cases is the same. Then the power ratio is (see equation 1)

$$\frac{P}{P'}=\left(\frac{G'}{G}\right)^2=\left(\frac{0.179}{0.197}\right)^2=0.825$$

i.e., 17.5% of the power can be saved by optimizing the current density distribution. From equation 15

$$J'=\frac{1}{\sqrt{v}}=0.100$$

Therefore, noting equation 14(A),

$$\frac{i_1}{i_1'}=\frac{J\sqrt{P}}{J'\sqrt{P'}}=\frac{JG'}{J'G}=\frac{0.226\times0.179}{0.100\times0.197}=2.05$$

The current density in the optimized coil at $r=a_1$ is 2.05 i_1', and decreases to 2.05 $i_1'\times0.206=0.424\ i_1'$ at $r=a_2$.

Example 2. Cylindrical Coil with Tapered Ends and Optimum Radial Current Distribution, $\rho^+=\lambda^+=1$; *see Fig. 4*

Equation 11(B) becomes in this case

$$i^+=\left(\frac{a_1}{r}\right)^2 \tag{20}$$

A cylindrical coil with tapered ends (resistivity and space factor constant) has a maximum Fabry factor when the current density is inversely proportional to the square of the distance from the axis. This is exactly the previously considered third example of section 3. For $k=1$ and $\alpha=3$, equation 10 yields $G_{\text{opt}}=0.204$. From equation 17, $J_{\text{opt}}=0.346$.

From equation 8 it follows that for a cylindrical coil with tapered ends, with $k=1$ and $\alpha=3$ but with constant current density, $G'=0.170$. With H_0 the same in each coil

$$\frac{P}{P'}=\left(\frac{0.170}{0.204}\right)^2=0.695$$

i.e., 30.5% power can be saved by optimizing the current distribution. From equation 15

$$J'=\frac{1}{\sqrt{\frac{4\pi}{3}(\alpha^3-1)}}=\frac{1}{\sqrt{108.9}}=0.0958$$

Therefore

$$\frac{i_1}{i_1'}=\frac{0.346\times0.170}{0.0458\times0.204}=3.01$$

The current density in the optimized coil at $r=a_1$ is 3.05 i_1' and decreases to $^1/_9$ of this value, i.e., to 0.334 i_1' at $r=a_2$. Whether or not the absolute value of the maximum current density is acceptable depends on the numerical value of the scaling factor

$$\sqrt{\frac{P}{\rho_1\lambda_1 a_1^3}}$$

7. Shape Optimization of Layer Coils

Section 4 dealt with finding the optimum current density distribution of a layer coil. Another problem is that of the optimum shape of a coil (i.e., where G is a maximum) when the current density distribution $i(r)$ is known. For simplicity, only symmetrical layer coils will be considered. The case of an unsymmetrical layer coil is discussed in Appendix II.

It can be shown (see Appendix I) that the optimization condition in this case is

$$\frac{\rho s^3 i}{r}=\frac{\rho(r^2+z^2)^{3/2} i}{r}=\text{constant} \tag{21}$$

Consider the following special cases:

Case 1: $\rho^+=i^+=1$, *Constant Current Density*

Using the co-ordinates s and ϕ (see Fig. 3), we obtain for the contour curve of the coil the equation

$$\frac{s^2}{\sin\phi}=\text{constant} \tag{22}$$

This case was considered by Fabry,[5] who found $G_{\max}=0.18$.

Case 2: $\rho^+=1$, $i^+=a_1/r$, *Current Density Inversely Proportional to the Distance from the Axis*

The equation of the contour curve is

$$\frac{s}{\sin^2\phi}=\text{constant} \tag{23}$$

The Fabry factor is again near 0.18.

Case 3: $\rho^+=1$, $i^+=(a_1/r)^2$, *Current Density Inversely Proportional to the Square of the Distance from the Axis*

In this case

$$\phi=\text{constant}=\frac{\pi}{4} \tag{24}$$

This layer coil is a coil with tapered ends, as considered in section 3, case 3, and section 6, example 2. It can be shown that this type of a layer coil represents the solution of the problem of finding the optimum Fabry factor when both shape and current density distribution are optimized simultaneously. However, in this case the maximum current density is relatively very high, and this combined with the difficulties of building such coils seems to limit their practical application.

8. Loop Coils

As mentioned previously, a loop coil, in contrast to a layer coil, is one in which the quantities i, r, ρ, and λ are functions of both co-ordinates r and z. Loop coils have little practical importance (although a slightly higher G is attainable) since their design is difficult. However, some basic theorems concerning loop coils and some numerical examples may be of interest.

Assume that a loop has an infinitesimally small cross-sectional area dA, in the z-r plane, and that its location is determined by the co-ordinates a and r (distance from the origin $s=\sqrt{z^2+r^2}$); then the magnetic field strength produced at the origin is (compare Bitter's computation of G for several special cases[7]):

$$H_0=\frac{2\pi}{10}\int_A \lambda\frac{r^2}{s^3}\,i\,dA \tag{25}$$

When normalized with respect to an appropriately selected reference loop (indicated by the subscript 1), this equation becomes

$$H_0=\frac{2\pi}{10}\lambda_1 i_1\int_A \lambda^+\frac{r^2}{s^3}i^+dA \tag{25A}$$

The power dissipated is

$$P=2\pi\int_A \rho\lambda r i^2 dA \tag{26}$$

or, normalized,

$$P=2\pi\rho_1\lambda_1 i_1^2\int_A \rho^+\lambda^+ r(i^+)^2\,dA \tag{26A}$$

Corresponding to equations 7(A) and 7(B) we obtain now for a loop coil

$$H_0=G\sqrt{\frac{P\lambda_1}{\rho_1 a_1}} \tag{27A}$$

$$G=\frac{\sqrt{2\pi a_1}}{10}\,\frac{\int_A \lambda^+\frac{r^2}{s^3}i^+\,dA}{\sqrt{\int_A \rho^+\lambda^+ r(i^+)^2\,dA}}\quad\text{(loop coil)} \tag{27B}$$

For the current density factor we obtain (compare equations 14(A) and 14(B)):

$$i_1=J\sqrt{\frac{P}{\rho_1\lambda_1 a_1^3}} \tag{28A}$$

$$J=\sqrt{\frac{a_1^3}{2\pi}}\,\frac{1}{\sqrt{\int_A \rho^+\lambda^+ r(i^+)^2 dA}}\quad\text{(loop coil)} \tag{28B}$$

Maxwell solved the problem of designing a galvanometer in which a given small electromotive force would produce the greatest possible deflection.[13] In a more general way we can ask about the optimum current density distribution of a loop coil of any shape, for which ρ and λ are known functions of z and r. Reasoning along the lines which led to finding the optimum current density distribution of a layer coil (see Appendix I) leads to the following optimization condition of a loop coil; compare with equations 11(A) and 11(B):

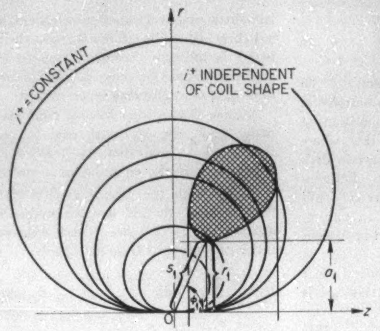


Fig. 5 (left). Optimum current density distribution of a loop coil

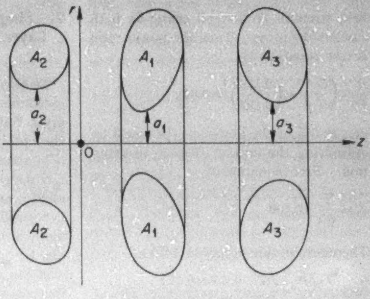


Fig. 7 (right). Set of optimized loop coils

$$\frac{\rho s^3 i}{r}=\frac{\rho_1 s_1^3 i_1}{r_1}=\text{constant} \tag{29A}$$

$$i^+=\frac{1}{\rho^+}\frac{s_1^3 r}{r_1 s^3}=\frac{1}{\rho^+}\frac{s_1^2}{\sin\phi_1}\frac{\sin\phi}{s^2} \tag{29B}$$

Substituting this value in equations 27(B) and 28(B) yields

$$G_{\text{opt}}=\frac{\sqrt{2\pi a_1}}{10}\sqrt{\int_A \frac{\lambda^+}{\rho^+}\left(\frac{\sin\phi}{s}\right)^2 dA} \tag{30}$$

$$J=_{\text{opt}}\frac{\sin^3\phi_1}{10G_{\text{opt}}} \quad \text{(loop coil)} \tag{31}$$

Equation 29(B) shows that the current density distribution does not depend on the shape of the cross-sectional area A of the coil, since the co-ordinates of the contour of the coil cross section do not appear in this equation. A family of curves can be drawn to represent the $i=$ constant lines which yield a maximum G for any shape of coil cross section; see Fig. 5. From equation 30 it follows that G_{opt} assumes the highest values if the coil cross section A extends from points as close as possible to its axis of symmetry. This is illustrated by the following examples of loop coils with optimized current density distribution.

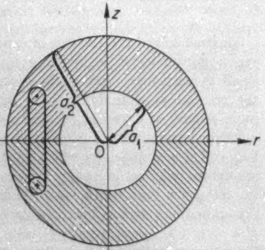


Fig. 6. Spherical coil

Example 1: Spherical Coil; Inside Radius a_1, Outside Radius a_2; see Fig. 6

A spherical coil is difficult to construct, since it must be built in two halves in order to provide access to the inner empty sphere. In this case

$$G_{\text{opt}}=\frac{1}{5}\sqrt{\frac{2\pi}{3}}\sqrt{\frac{\alpha-1}{\alpha}} \tag{32}$$

For very large α the Fabry factor approaches the value 0.289.[5]

Example 2: Long Cylindrical Solenoid; Inside Radius a_1, Outside Radius a_2

$$G_{\text{opt}}=\frac{\pi\sqrt{3}}{20}\sqrt{\frac{\alpha-1}{\alpha}} \tag{33}$$

For very large α the Fabry factor approaches the value 0.272.[6,7]

Example 3: Cylindrical Coil With Tapered Ends (Fig. 4)

$$G_{\text{opt}}=\frac{\sqrt{2\pi}}{10}\sqrt{\frac{3}{4}\left(\frac{\pi}{2}-\phi\right)+\frac{1}{2}\sin 2\phi-\frac{1}{16}\sin 4\phi}\;\sqrt{\frac{\alpha-1}{\alpha}} \tag{34}$$

For $\phi=\pi/4$

$$G_{\text{opt}}=0.262\sqrt{\frac{\alpha-1}{\alpha}} \tag{34A}$$

This yields $G_{\text{opt}}=0.213$ for $\alpha=3$.

As shown in example 2 of section 6, cylindrical layer coil of the same shape optimized in respect to current density distribution has a Fabry factor $G_{\text{opt}}=0.204$, which is only slightly smaller. The current density factor J_{opt}, however, increases noticeably from 0.346 for the layer coil to 0.470 for the optimized loop coil. This shows clearly the merely academic interest of this case.

Finally, the following theorem might be derived. Consider a set of n loop coils of any cross sections; see Fig. 7. The optimized Fabry factor of the coil k is, from equation 30,

$$G_{k\,\text{opt}}=\frac{\sqrt{2\pi a_k}}{10}\sqrt{\int_{A_k}\frac{\lambda^+}{\rho^+}\left(\frac{\sin\phi}{s}\right)^2 dA} \tag{35}$$

The optimized Fabry factor of the whole loop arrangement is

$$G_{\text{opt}}=\frac{\sqrt{2\pi a_1}}{10}\sqrt{\int_A\frac{\lambda^+}{\rho^+}\left(\frac{\sin\phi}{s}\right)^2 dA} \tag{36A}$$

The integration has to be performed over the total cross-sectional area A, i.e., over all the incremental areas A_k. Expressed in somewhat different notation,

$$G_{\text{opt}}=\frac{\sqrt{2\pi a_1}}{10}\sqrt{\sum_{k=1}^{n}\int_{A_k}\frac{\lambda^+}{\rho^+}\left(\frac{\sin\phi}{s}\right)^2 dA_k} \tag{36B}$$

Therefore

$$\frac{G_{\text{opt}}^2}{a_1}=\sum_{k=1}^{n}\frac{G_{k\,\text{opt}}^2}{a_k} \tag{37}$$

The Fabry factor is increased by the addition of more coils to the set. Its highest value is achieved when the whole available space is filled with windings. This is in contrast to layer coils. For these the Fabry factor decreases when a certain optimum winding space is exceeded.

Conclusions

1. General expressions for the Fabry factor of a layer coil (for which coil variables depend on the distance r from the axis only), and for a loop coil (coil variables depend on r and on axial coordinate z) have been derived; see equations 7(B) and 27(B). These general forms are convenient for the computation of the Fabry factor of a coil of any shape and any distribution of current density, resistivity, and space factor.

2. To produce a desired field strength

H_0 at any point on the axis, with the minimum power P, the following kinds of optimization are possible:

The coil shape and the distribution of ρ and λ are known; the problem is to find the optimum current density distribution. For symmetrical loop coils and layer coils with variable ρ and λ, general solutions have been found; see equations 11(B) and 29(B). The following result is of particular interest: A desired magnetic field strength in the center of a symmetrical layer coil is achieved with the minimum power when the product of resistivity ρ, radial distance r of a layer from the axis, distance of the layer end point from the coil center s, and current density i is constant.

The current density distribution of a layer coil is any known function $i=i(r)$, likewise the distributions of ρ and λ are known; find the shape of the optimum coil. A general solution has been found; see equation 21. Of particular interest is the current density distribution

$$i=\left(\frac{a_1}{r}\right)^2 i_1$$

In this case the optimum coil is cylindrical with tapered ends.

Simultaneously optimizing current density distribution and coil shape likewise leads to the cylindrical coil with tapered ends and current density inversely proportional to the square of the distance from the axis.

3. For any magnet coil, the current density i_1 in the innermost layer (or in the loop with smallest radius) can be represented by the product of a dimensionless factor, the current density factor J, and a scaling factor

$$i_1=J\sqrt{\frac{P}{\rho_1\lambda_1 a_1^3}} \qquad (14A)$$

The analogy to the Fabry factor is obvious. General expressions for the current density factors of layer and loop coils have been derived [equations 14(B) and 28(B)]. For a coil of any shape with constant resistivity, space factor, and current density, the current density factor J is the reciprocal value of the square root of the volume factor; see equation 15.

4. For coils with optimum current density distribution, simple general expressions for the Fabry factor G_{opt} have been derived; see equations 12 and 30. In this case the current density factor of a layer coil becomes

$$J_{\text{opt}}=\frac{a_1}{10 s_1}\frac{1}{G_{\text{opt}}}\text{(layer coil)} \qquad (17)$$

and that of a loop coil

$$J_{\text{opt}}=\frac{\sin^2\phi_1}{10}\frac{1}{G_{\text{opt}}}\text{(loop coil)} \qquad (31)$$

5. The Fabry factor, optimized with respect to current density, of n loop coils, each having the optimized Fabry factors $G_{k\,\text{opt}}$, is

$$\frac{G_{\text{opt}}^2}{a_1}=\sum_{k=1}^{n}\frac{G_{k\,\text{opt}}^2}{a_k} \qquad (37)$$

Numerical examples illustrate the basic concepts and theorems mentioned above.

Appendix I

Bitter[7] showed how to find the optimum current density distribution and the corresponding values of the Fabry factor of coils of known shape. He did this by the method of indeterminate multipliers, and the same mathematical method shall be applied here to the optimization of the current density $i(r)$ of a layer coil of any symmetrical shape and with any $\rho(r)$ and $\lambda(r)$ distribution.

From equations 4 and 5

$$F[i(r)]=H_0+\mu P=\frac{4\pi}{10}\int_{r=a_1}^{a_2}\lambda i\frac{z}{s}dr+\mu 4\pi\int_{r=a_1}^{a_2}\rho\lambda i^2 rz\,dr \qquad (38)$$

μ being an arbitrary constant.

$$\delta F[i(r)]=0 \qquad (39)$$

is satisfied if

$$\frac{4\pi}{10}\lambda\frac{z}{s}+\mu 8\pi\rho\lambda irz=0 \qquad (40)$$

Therefore

$$\rho rsi=-\frac{1}{20\mu}=\text{constant}=\rho_1 a_1 s_1 i_1 \qquad (41)$$

This is equation 11(A). Note that $\lambda(r)$ does not occur in this optimization condition for $i(r)$. The space factor $\lambda(r)$, of course, influences the Fabry factor, as shown in equation 12. A similar calculation can be used to prove equation 21.

Appendix II

Equations for Asymmetrical Layer Coils

Fig. 8 shows a layer coil with the z-axis as the axis of rotational symmetry. The coil, however, is not symmetrical to any plane perpendicular to the z-axis (asymmetrical layer coil). Referring to Fig. 8, the abbreviation

$$\cos\phi_{II}-\cos\phi_I=\psi \qquad (42)$$

is used. The magnetic field strength at the point 0 assumes the form

$$H_0=\frac{\pi}{5}\int_{r=a_1}^{a_2}\lambda\psi i\,dr \qquad (43)$$

This can be written in normalized form:

$$H_0=\frac{\pi}{5}\lambda_1 i_1\int_{r=a_1}^{a_2}\lambda^+\psi i^+dr \qquad (44)$$

The ohmic power loss becomes

$$P=2\pi\int_{r=a_1}^{a_2}\rho\lambda r(z_{II}-z_I)i^2dr$$
$$=2\pi\rho_1\lambda_1 i_1^2\int_{r=a_1}^{a_2}\rho^+\lambda^+r(z_{II}-z_I)(i^+)^2dr \qquad (45)$$

From these equations it follows that the Fabry factor

$$G=\frac{1}{5}\sqrt{\frac{\pi a_1}{2}}\frac{\int_{r=a_1}^{a_2}\lambda^+\psi i^+dr}{\sqrt{\int_{r=a_1}^{a_2}\rho^+\lambda^+r(z_{II}-z_I)(i^+)^2dr}} \qquad (46)$$

and the current density factor

$$J=\sqrt{\frac{a_1^3}{2\pi}}\frac{1}{\sqrt{\int_{r=a_1}^{a_2}\rho^+\lambda^+r(z_{II}-z_I)(i^+)^2dr}} \qquad (47)$$

The current density distribution for which G is a maximum can be found by the method discussed in Appendix I. The result is

$$\frac{\rho(z_{II}-z_I)r}{\psi}i=\text{constant} \qquad (48)$$

If we consider a coil cross section similar to that shown in Fig. 7, equation 48 applied to the point $r=a_1$ assumes the form 0/0, which evaluates to

$$\frac{\rho(z_{II}-r_I)r}{\psi}i=\frac{\rho_1 a_1^2}{\sin^2\phi_1}i_1 \qquad (49)$$

or

$$i^+=\frac{\psi}{\sin^2\phi_1}\frac{a_1^2}{(z_{II}-z_I)r\rho^+} \qquad (50)$$

This value substituted in equations 46 and 47 leads to

$$G_{\text{opt}}=\frac{1}{5}\sqrt{\frac{\pi a_1}{2}}\sqrt{\int_{a_1}^{a_2}\frac{\lambda^+\psi^2}{\rho^+}\frac{dr}{(z_{II}-z_I)r}} \qquad (51)$$

$$J_{\text{opt}}=\frac{\sin^2\phi_1}{10}\frac{1}{G_{\text{opt}}} \qquad (52)$$

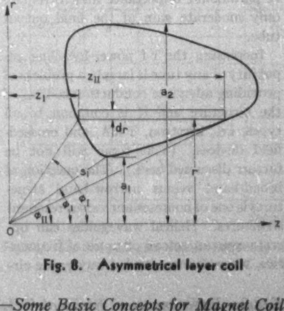


Fig. 8. Asymmetrical layer coil

References

1. PROJECT SHERWOOD. THE U. S. PROGRAM IN CONTROLLED FUSION (book), A. S. Bishop. Addison-Wesley, Reading, Mass., 1958.

2. PROCEEDINGS OF SYMPOSIUM ON MAGNETIC FIELD DESIGN IN THERMONUCLEAR RESEARCH. *ORNL-2745*, Offices of Technical Services, Department of Commerce, Washington, D. C., Dec. 1958.

3. THE ORNL THERMONUCLEAR PROGRAM, P. R. Bell, et al. *AEC Research and Development Report, ORNL-2745*, Oak Ridge National Laboratories Oak Ridge, Tenn., Jan. 15, 1958.

4. SUR LE CHAMP MAGNETIQUE AU CENTRE D'UNE BOBINE CYLINDRIQUE, C. Fabry. *L'Eclairage Electrique*, Paris, France, vol. XVII, no. 43, 1898, pp. 133–41.

5. Production de Champs Magnétiques Intenses au Moyen de Bobines sans Fer, C. Fabry. *Journal de Physique*, Paris, France, vol. IX, 1910, pp. 129–34.

6. The Design of Coils for the Production of Strong Magnetic Fields, J. D. Cockcroft. *Proceedings*, Royal Society, London, England, vol. CCXXVII, no. A654, 1928, pp. 317–43.

7. The Design of Powerful Electromagnets, F. Bitter. *Review of Scientific Instruments*, New York, N. Y., vol. VII, 1936, pp. 479–87; vol. VII, 1937, pp. 318–19; vol. IX, 1939, pp. 373–81.

8. See reference 4, fig. 2.

9. Magnetic Field Design of Mirror Coils, W. F. Gauster. *TID-7558*, Technical Information Service Report, Washington, D. C., Feb. 3–5, 1958, pp. 518–27.

10. See reference 4, p. 138.

11. Electric and Magnetic Fields (book), S. S. Atwood. John Wiley & Sons, Inc., New York, N. Y., 1939.

12. Coils Without Iron for the Production of Very Intense Continuous Magnetic Fields, F. Gaume. *Journal des Recherches*, Centre National de la Recherche Scientifique, Laboratoires de Bellevue, Paris, France, vol. 43, June 1958. (Translated into English by H. H. Kolm. *Group Report M81-12*, Lincoln Laboratory, Lexington, Mass., Sept. 17, 1959.)

13. A Treatise on Electricity and Magnetism, J. C. Maxwell. Oxford University Press, Oxford, England, third edition, 1892, pp. 360–64.

Experiments on a Series of S-Band Crestatrons

J. E. ROWE
ASSOCIATE MEMBER AIEE

G. T. KONRAD
ASSOCIATE MEMBER AIEE

H. W. KRAGE
NONMEMBER AIEE

IN RECENT YEARS, power klystrons, traveling-wave amplifiers (TWA), and various forms of crossed-field amplifiers and oscillators have been produced in relatively large numbers with essentially reproducible characteristics. Gradually these devices have found their way into more and more sophisticated systems applications with the logical result that they are competing with one another with regard to the most desirable set of characteristics. This competition deals with weight, efficiency, bandwidth, average power capability, and modulation characteristics, to name just a few items. In general it is not feasible to optimize or maximize all of these parameters in one tube and hence there seems to be room for all the above-mentioned device types, depending upon the particular application and which of the characteristics is most critical. The day of the high-gain high-power high-efficiency broad-bandwidth short-length lightweight low-noise amplifier to be powered by a solar cell is not yet here. The Crestatron amplifier to be described in this paper affords a means of optimizing several of these characteristics in one device.

The conventional growing-wave TWA utilizing a helical line as an r-f (radio-frequency) structure has a broad bandwidth capability due to the nonresonant nature of the helix. The helix is incapable of dissipating much r-f power because it in general can be radiation cooled only when supported inside a glass vacuum envelope. For example, an S-band TWA cannot handle more than approximately 250 watts of r-f power if only radiation cooling is used. Some increase has been obtained by using conduction cooling of the helix through ceramic rods to an outside metallic vacuum envelope. R-f structures that can handle more power are usually resonant periodic waveguides and these have a more limited bandwidth than the helix. Both types of TWA require an r-f attenuator capable of dissipating the output power in order to prevent oscillations due to reflections when the gain is high.

Unfortunately this attenuator limits the efficiency and power output and, of course, reduces the gain; hence it makes the tube longer for any given gain as compared with an attenuatorless tube.

In transmitter output tubes where power supply requirements are severe and high efficiency is very desirable, it would seem that a power tube with high efficiency, broad bandwidth, short length, and no attenuator would be most desirable in almost any systems application. It is feasible to build high gain in the low-power stages where high efficiency is not of paramount importance and to require only moderate gain of the final output tube.

Increasing the r-f power-handling capability of any tube is largely a problem of providing adequate conduction cooling of the structure and it is common to all types, i.e., klystron, TWA, and crossed-field devices. (This topic will not be further discussed here.) The question of broad-band versus narrow-band structures is one of nonresonant versus resonant structures. Helical waveguides can operate over an octave or more of frequencies, whereas most loaded-waveguide circuits operate only over a bandwidth of 5% to 15% of the center frequency.

In recent years, two devices operating at moderate gains without attenuators and at high efficiencies have been developed. One is a crossed-field device called the Amplitron and the other is an O-type device called the Crestatron.[1,2] The Amplitron may be described as an injection-locked oscillator having a nominal gain of 10 db (decibels), a bandwidth of 10%, and an operating efficiency of some 60%; it will not be further discussed here. The object of this paper is to present experimental results obtained on a series of S-band Crestatron amplifiers.

Crestatron Principle

The Crestatron amplifier is a form of traveling-wave device which operates on a beating-wave principle in that three constant-amplitude r-f waves traveling at different phase velocities along an r-f circuit beat with one another to produce a power gain. Two of the waves are out of phase at the input and the device length is chosen so that at its output the waves add in phase. It will be shown later in this paper that this is a broad-band

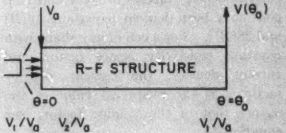


Fig. 1. Illustration of beating-wave amplification

Paper 60-1411, recommended by the AIEE Electronics Committee and approved by the AIEE Technical Operations Department for publication. Manuscript submitted February 15, 1960; made available for printing September 29, 1960.

J. E. Rowe, G. T. Konrad, and H. W. Krage are with the University of Michigan, Ann Arbor, Mich.

Vol.
Pt. 1
196

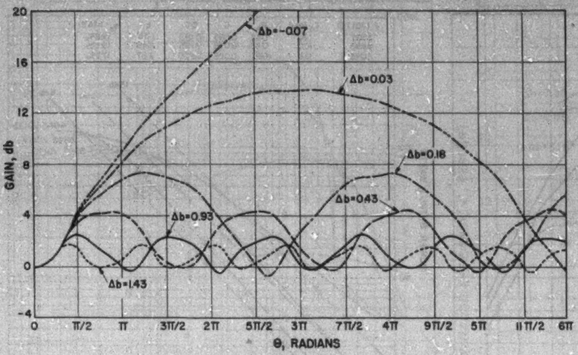


Fig. 2. Gain versus length; $C=0.1$, $QC=0.25$, $d=0$, $b_{x_1=0}=2.57$

phenomenon. A solution of the secular equation describing the effect of the beam on the circuit and the effect of the circuit on the stream gives a third-order equation in the propagation constants. The three waves result when an r-f voltage is applied to the input of the amplifier where the electron stream enters the interaction region, and this splits into three waves when the continuity of electron stream velocity and convection current are satisfied across the input boundary. This beating-wave phenomenon is illustrated in Fig. 1.

The Crestatron is operated with the electron stream velocity sufficiently larger than the circuit phase velocity so that the wave propagation constants are purely imaginary and only constant-amplitude waves traveling at different phase velocities exist on the structure. This stream voltage (velocity) is considerably higher than that used in the conventional growing-wave amplifier. It leads to an increased efficiency and power output over the growing-wave device. The gain of the Crestatron is proportional to the excess velocity over that at which the conventional growing-wave gain goes to zero. This is the point on the propagation constant diagram where the propagation constants change from complex values to purely imaginary ones. The detailed theory behind the Crestatron is given in reference 2.

Let us define the value of the velocity parameter b for which the growing-wave constant x_1 goes to zero as $b_{x_1}=0$ and then define Δb as the excess over this value.

$$b \triangleq \frac{1}{C}\left(\frac{u_0-v_0}{v_0}\right) \tag{1}$$

and

$$\Delta b \triangleq b-b_{x_1=0} \tag{2}$$

where

C = gain parameter
$u_0=(2\eta V_0)^{1/2}$, electron stream velocity
v_0 = circuit phase velocity

The gain of the Crestatron for a typical set of circuit and stream parameters is shown in Fig. 2. The parameters used in the figure are:

QC = space-charge parameter
$\theta=2\pi CN_s$, radian tube length
N_s = length of tube, stream wavelengths

The gain is calculated from the following voltage ratio, which involves the initial wave amplitudes and the propagation constants:

$$\frac{V(\theta)}{V_c(0)}=e^{-j\frac{\theta}{C}}\sum_{i=1}^{3}\left(\frac{V_{ci}}{V_c(0)}\right)e^{\delta_i\theta} \tag{3}$$

These gain calculations were made with a linear theory and it is seen that as Δb increases, the gain decreases and the device becomes shorter. The optimum length is seen to be approximately six slow wavelengths for the case shown.

For shortest device length it is always desirable to operate at the first maximum of the gain curve. This optimum length is given below as a function of Δb.

$$CN_s=\frac{\theta}{2\pi}=\frac{0.25}{(\Delta b)^{1/2}} \tag{4}$$

As the gain parameter C is increased through an increase in either the stream perveance or the circuit impedance, the optimum length decreases and the gain increases. The highest value of C achievable depends upon the frequency and the type of circuit used. Either a helical waveguide or any form of periodically loaded waveguide may be utilized. If C is considered constant the abscissa of Fig. 2 is proportional to frequency and it can be seen that the 3-db bandwidth is approximately 3 to 1. Clearly, then, this is a broad-band phenomenon.

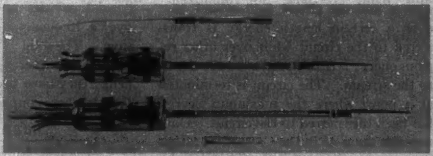


Fig. 3. S-band 100-watt Crestatron (approximately 1/5 size); $G=10$–12 db, $\eta>20\%$

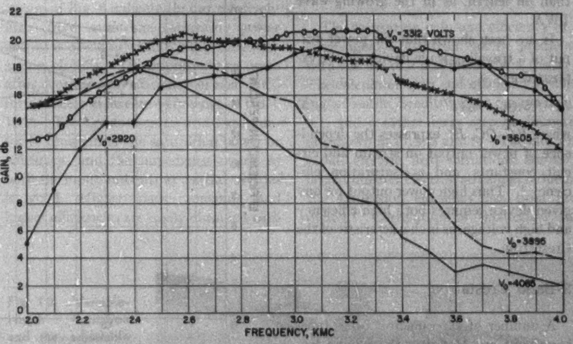
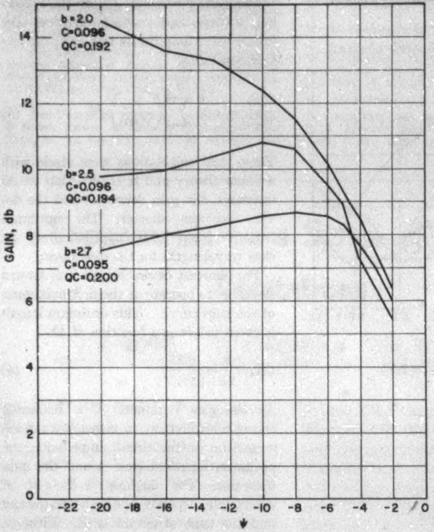


Fig. 4. Small-signal gain versus frequency for S-8-3 Crestatron. Drive level≅100 milliwatts; at 3.0 kmc, $C=0.095$, $QC=0.185$, $P\mu\cong0.60$



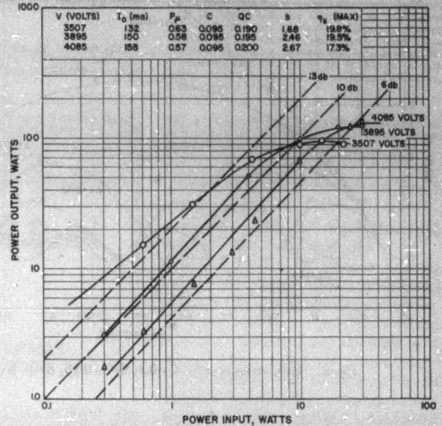


Fig. 5 (left). Saturation gain versus ψ (input signal level in db relative to $CI_0 V_0$) with b as parameter for S-8-3 Crestatron at 3.0 kmc; d = 0.0279

Fig. 6 (above). Power output versus power input for the S-8-3 Crestatron at 3 kmc

There is power gain in the device, and the added energy at the output comes from the electron stream. The three waves set up on the structure travel along at different phase velocities and the phase relationship between the r-f current in the stream and the circuit field is such that over a certain length energy is transferred to the circuit while over other regions energy is transferred from the circuit to the stream. The circuit is terminated at a point where the maximum energy has been transferred to the circuit. Gain depends upon Δb in the Crestatron rather than on length, as in the growing-wave TWA.

It has been shown that the power output of a traveling-wave device can be expressed as

$$P_0 = f(C,QC,B)\eta_s V_0^{5/2} \tag{5}$$

where $f(C, QC, B)$ expresses the dependence of power output on stream and circuit constants, and η_s = saturation efficiency.[3] Thus high power output for any given device requires both high efficiency and high voltage, as is characteristic of the Crestatron.

S-Band Crestatron

A number of experimental tubes have been built and tested in the S-band frequency range. The power level, efficiency, and gain are generally functions of the electron stream perveance. A typical S-band model giving over 120 watts is shown in Fig. 3. The active circuit (helix) is 2.5 inches and 3.75 inches long in the two tubes shown. The input and output transducers used were coaxial cavities and the tube operated over a frequency range of nearly an octave. The solid cylinders at the ends of the helix are necessary to terminate the helix; they have been eliminated on later models where coupled-helix couplers were used. This, of course, results in a further reduction in the tube length. The helix in this tube was not specifically designed for Crestatron operation but was taken from a growing-wave tube and simply made shorter in accordance with small-signal calculations. The electron gun was of modest perveance and the tube uses electromagnet focusing.

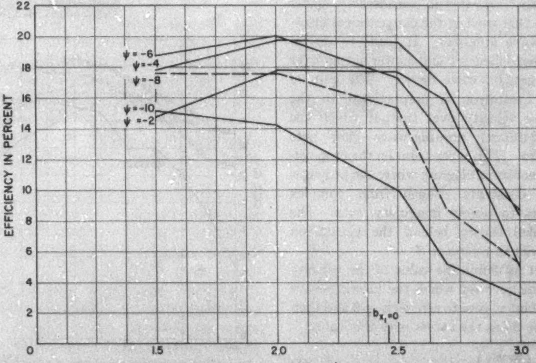


Fig. 7. Saturation efficiency versus b with ψ as parameter for S-8-3 Crestatron; C≅0.095, QC≅0.20, d≅0.028, $b_{x_1=0}$≅2.46

The small-signal gain of this S-band Crestatron is shown in Fig. 4 as a function of frequency for various values of stream

Vol.
Pt. 1
196

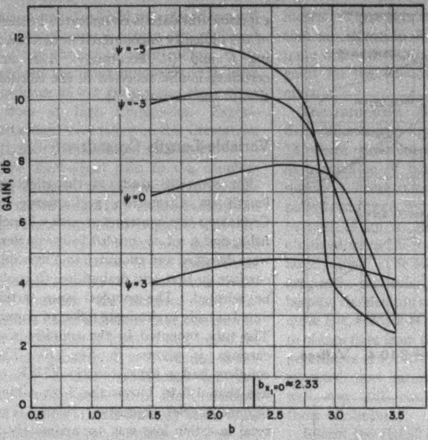


Fig. 8 (above). Saturation gain versus b for constant drive powers, C=0.1, QC=0.125, d=0, B=1

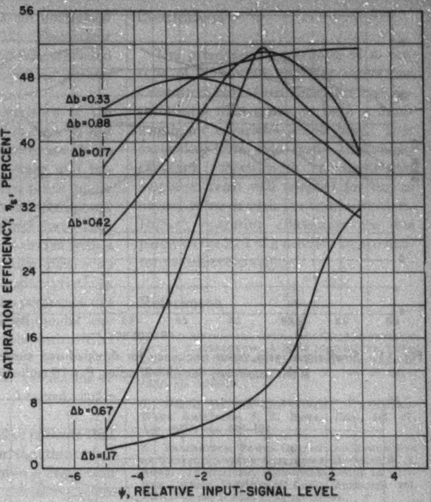


Fig. 9 (right). Efficiency versus drive power for constant b, C=0.1, QC=0.125, d=0, B=1

voltage (electron velocity). The high gain is associated with the fact that the drive level is small: 100 milliwatts. It is seen that the 3-db bandwidth is nearly an octave. The two reasons for the bandwidth limitation are: 1. The transducers have a high voltage standing-wave ratio (VSWR) at the band edges. 2. The helix is dispersive, giving rise to a higher impedance at the low frequencies, which results in a slower drop-off of the gain at high voltages for low frequencies. The coupler VSWR is less than 1.4 to 1 over most of the band. The total insertion loss including helix and both couplers is approximately 2 db.

The gain-versus-frequency characteristic at high voltages can easily be flattened by placing a metallic shield over the helix at a radius of 1.5 to 2 times the mean helix radius. This has the effect of lowering the low-frequency impedance of the helix more than the high-frequency impedance since the fields extend further from the circuit at low frequencies than at high frequencies. The shield tends to divert a fraction of the helix field to the region between the helix and the shield where it is not seen by the electrons. At low frequencies the gain is due to a growing-wave interaction, since x_1 is not zero here. This indicates that an extremely broad-band operation can be obtained by designing the tube to operate as a Crestatron in the high-frequency ranges and operate as a growing-wave amplifier at low frequencies. It is possible to effect a design of this type and obtain a flat gain curve versus frequency over a wide range.

The gain as a function of the drive level to the amplifier is shown in Fig. 5 for a range of some 20 db of drive relative to the product of the gain parameter times the d-c beam power CI_0V_0. For high voltages and moderate gains the saturation gain is relatively independent of the drive level, which is sometimes a desirable limiting characteristic. At high values of the amplifier drive power, even though the saturation output increases, the gain decreases to some 3 to 6 db.

The power output versus power input curves of Fig. 6 indicate that at relatively high drive levels the saturation power output is near 130 watts for this tube with a gain of 6 db. This represents an operating (conversion) efficiency of near 20% for an electron stream perveance of 0.6×10^{-6}. The Crestatron has no r-f attenuator to limit either the saturation power or efficiency, and the impedance match of both the input and output coaxial cavities is good. Higher power outputs and higher efficiencies are easily obtained with higher electron stream perveances. A stream perveance of 1×10^{-6} will give saturation efficiencies of well over 30%. It is seen that saturation gains of 10 db with a power output of 100 watts can be obtained simultaneously.

In the beating-wave regime of operation the saturation efficiency generally increases with the drive level, reaching a maximum for relatively high input powers, as shown in Fig. 7. Generally low values of the space-charge parameter, $QC \approx 0.1-0.2$, are desirable in order to achieve highest conversion efficiencies and power outputs. Circuit loss tends to reduce both the gain and the efficiency.

Theoretical Predictions

The saturation gain and efficiency of a large-signal Crestatron can be predicted with the use of a nonlinear TWA theory.[4] The calculated gain and efficiency assuming a lossless circuit, a gain parameter of 0.1, and a rather large stream diameter are shown in Figs. 8 and 9. The effect of radial electric field variations across a large-diameter electron stream in reducing the saturation efficiency has been predicted.[4] Correlation of theoretical and

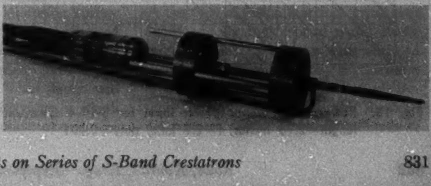


Fig. 10. Variable-length Crestatron and the adjustable coupled-helix assembly

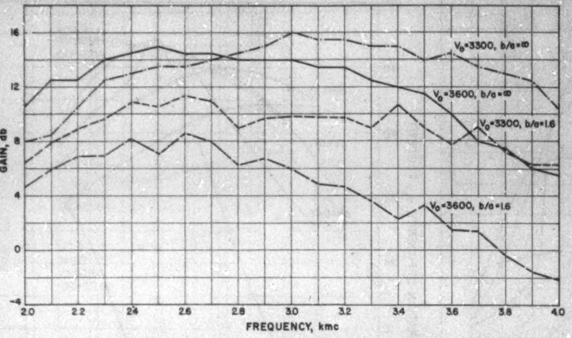


Fig. 11. Small-signal gain versus frequency for shielded and unshielded S-10-4. Voltage is the parameter; length = 3 inches, P_i = 10 milliwatts

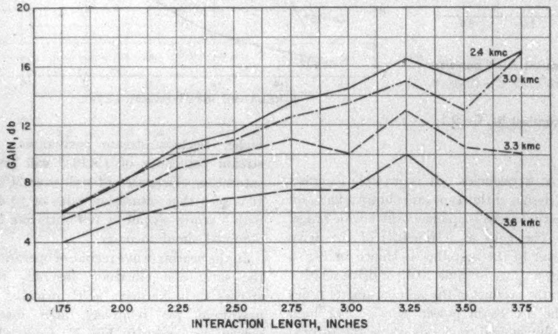


Fig. 12. Small-signal gain versus interaction length for S-10-4 Crestatron; V_0 = 3,700, $b/a = \infty$

experimental data is complicated because of the difficulty encountered in determining C and QC accurately. This uncertainty is also reflected in the calculation of the operating Δb.

Variable-Length Crestatron

For complete study of the effect of length on Crestatron performance, a Crestatron was constructed with a 7-inch helix, and a set of coupled-helix couplers were designed and mounted in a movable carriage so that any desired length could be selected. The couplers were terminated outside to eliminate reflected power. The tube mounted in the couplers and carriage is shown in Fig. 10. The couplers had a characteristic VSWR of less than 1.6 to 1 over the 2- to 4-kmc (kilomegacycle) frequency range. The total insertion loss was approximately 2 db for the two couplers and the interaction helix. The same helix design was used in this variable-length Crestatron as in the previously discussed tube, the *S-10-4*, which was operated without an attenuator as was the *S-8-3*, and like its predecessor utilized an electromagnet for focusing of the stream.

On this Crestatron, with the use of a relatively low-perveance electron gun, a vast amount of data has been taken, but only a small part will be presented here. It is planned in the future to carry out similar tests using higher-perveance guns. As mentioned earlier, the helix used was slightly dispersive at low frequencies and

Vol.
Pt. 1
1960

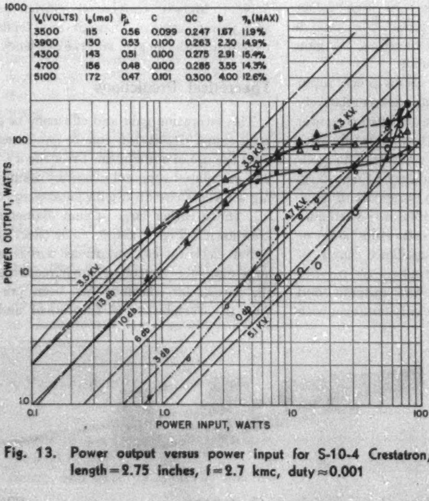

V_0(VOLTS)	I_0(ma)	P_μ	C	QC	b	η_e(MAX)
3500	115	0.56	0.099	0.247	1.67	11.9%
3900	130	0.53	0.100	0.263	2.30	14.9%
4300	143	0.51	0.100	0.275	2.91	15.4%
4700	156	0.48	0.100	0.285	3.55	14.3%
5100	172	0.47	0.101	0.300	4.00	12.6%

Fig. 13. Power output versus power input for S-10-4 Crestatron; length = 2.75 inches, f = 2.7 kmc, duty ≈ 0.001

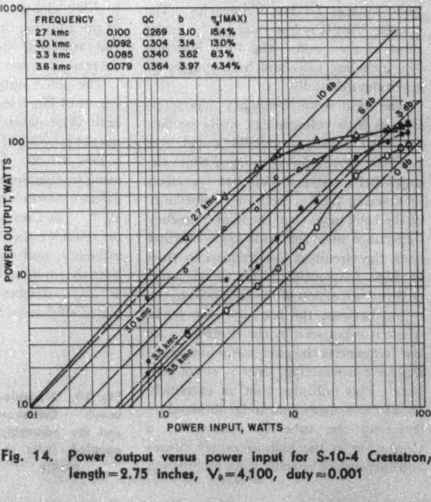

FREQUENCY	C	QC	b	η_e(MAX)
2.7 kmc	0.100	0.269	3.10	15.4%
3.0 kmc	0.092	0.304	3.14	13.0%
3.3 kmc	0.085	0.340	3.62	8.3%
3.6 kmc	0.079	0.364	3.97	4.34%

Fig. 14. Power output versus power input for S-10-4 Crestatron; length = 2.75 inches, V_0 = 4,100, duty ≈ 0.001

hence a metallic shield was placed around the glass vacuum envelope at a radius of 1.6 times the inner helix radius. This shield has the effect of lowering the helix impedance at low frequencies more than it does at high frequencies. Experimental data were taken on this Crestatron both with the shield and without it.

The small-signal gain as the coupled-helix couplers were adjusted to give an interaction length of 3 inches, is shown in Fig. 11 as a function of frequency and voltage both for the shielded and unshielded cases. A comparison of the gain-versus-frequency results for particular voltages between the shielded and unshielded cases indicates that the gain at low frequencies is lowered and hence the constant-gain region is broadened in frequency. Octave bandwidths are evidently easily achieved in the Crestatron and this limitation is due only to the finite bandwidths of the couplers.

It is interesting to plot the gain versus the interaction length for this Crestatron at particular frequencies since it shows that the optimum length from both a wide bandwidth and a reasonable gain standpoint is approximately 3 to 3.25 inches. These results for the unshielded case are shown in Fig. 12. They agree quite well with the theoretically predicted gains for the various operating parameters.

The saturation characteristics of the *S-10-4* Crestatron appear in Figs. 13 and 14, where power output curves are shown as a function of both voltage and frequency. These data were taken under pulsed conditions and it is seen that power outputs of up to 190 watts have been obtained when the tube is driven with a large signal. The saturation efficiencies are not particularly high, primarily because the electron stream perveance was quite low. Work is presently under way to obtain data from this tube with stream perveances of 1×10^{-6} and 2×10^{-6}. It is expected that saturation efficiencies of well over 30% can be obtained under these conditions.

Recent theoretical studies indicate that a significant increase in the theoretical efficiency can be obtained in a Crestatron by prebunching of the electron stream. This also has the effect of increasing the gain and further reducing the device length.

Conclusions

The principles of operation of a new type of microwave amplifier called the Crestatron have been described and experimental data given on both a fixed-length and a variable-length tube. It is concluded from this study that high-efficiency short-length moderate-gain amplifiers seem well adapted for use as output tubes in high-power transmitters. Highest operating efficiencies are due partly to the lack of a need for an attenuator to achieve stability.

References

1. Description and Operating Characteristics of the Platinotron—A New Microwave Tube Device, W. C. Brown. *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 45, Sept. 1957, pp. 1209–23.

2. Theory of the Crestatron: A Forward-Wave Amplifier, J. E. Rowe. *Ibid.*, vol. 47, Apr. 1959, pp. 536–45.

3. Theoretical Power Output and Bandwidth of Traveling-Wave Amplifiers, H. Sobol, J. E. Rowe. *Transactions*, Professional Group on Electron Devices, Institute of Radio Engineers, vol. PGED–7, no. 2, Apr. 1960, pp. 84–94.

4. One Dimensional Traveling-Wave Tube Analyses and the Effect of Radial Electric Field Variations, J. E. Rowe. *Ibid.*, no. 1, Jan. 1960, pp. 16–21.

Fusion Power from Magnetically Confined Plasma

R. G. MILLS
MEMBER AIEE

ON THE BASIS of an assumed magnetically confined, stable, quiescent plasma, the relations for the fusion power liberated and the electric power consumed to generate an externally applied magnetic field are recalled. From these, appropriate scaling laws are developed which show the relative importance of various machine parameters, independently of the specific form of machine assumed. In the event that any of this class of machines should be proved physically possible, the directions of development needed to improve the economy of a resulting power-generating system are indicated. The advantages and disadvantages of various methods of increasing the yield of a fusion plant are compared.

Release of Energy by Thermonuclear Fusion

For the purposes of this paper, it is necessary to assume that a successful method will be found to confine a thermonuclear plasma in a quiescent, stable manner for periods long enough to allow usefully large amounts of fusion energy to be liberated. It is further assumed that the basic confinement tool is an externally applied magnetic field of an unspecified geometry. This paper will indicate some of the directions in which development could be undertaken to increase the economy of any fusion reactor resulting from this postulated successful magnetic confinement method.

The basic physics of fusion reactions has been reviewed in a paper by Post,[1] in which it is shown that the fusion power density of a high-temperature gas of two fusible isotopes is given by

$$p_F = n_1 n_2 (\overline{\sigma v}) E \tag{1}$$

where n_1 and n_2 are the particle densities of the two reacting isotopes, $(\overline{\sigma v})$ is the product of the reaction cross section and the relative velocity of two reacting particles averaged over the velocity distribution of the gas, and E is the energy liberated in a single fusion event. In the case of a reacting gas of a single species such as deuterium, the product $n_1 n_2$ is replaced by $n^2/2$.

In an operating reactor, whatever may be the fuel, a certain mixture will be established and a certain average energy E will be produced per fusion event. The exact value of the number E will depend on the type of fusion reactions being produced and by the disposal made of the energy carried by the resulting neutrons such as by thermalization, regenerative production of fuel, production of isotopes, or production of fission. We need not concern ourselves with the details of

---

Paper 60-1410, recommended by the AIEE Nucleonics Committee and approved by the AIEE Technical Operations Department for publication. Manuscript submitted April 15, 1960; made available for printing September 29, 1960.

R. G. Mills is with Project Matterhorn, Princeton University, Princeton, N. J.

This paper was presented at the AIEE Symposium on the Engineering Aspects of Magnetohydrodynamics, at the University of Pennsylvania, February 19, 1960. The work was supported by the Atomic Energy Commission under Contract no. *AT(30-1)-1238*.

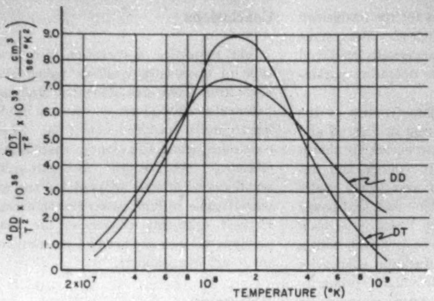


Fig. 1. The product $\overline{\sigma v}/T^2$ as a function of temperature

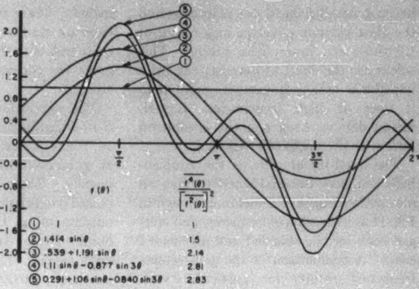


Fig. 2. Waveshapes maximizing the ratio of $\overline{B^4}$ to $(\overline{B^2})^2$

the energy budget, but merely note that the fusion power density, when the constants are dropped, is given by the proportion

$$p_F \sim n^2(\overline{\sigma v}) \tag{2}$$

An important dimensionless parameter in plasma physics has been given the name β. This represents the ratio of gas pressure to external magnetic field energy density, or

$$\beta \equiv \frac{8\pi nkT}{B^2} \tag{3}$$

If a magnetic field is used to confine a plasma, this number must be less than 1, since $\beta = 1$ represents the situation where the gas pressure and the effective magnetic pressure are equal and any slight increase in material pressure will cause the gas to burst through the confinement. In practice, however, β is likely to be limited to values far less than 1 (perhaps 3% to 10%) because of stability considerations. The theory of hydromagnetic stability[2] indicates that for any given confinement geometry there exists a critical β above which the system will become unstable. Thus β is an important parameter, and it is convenient to use equation 3 to solve for the density n, in terms of the other variables, or

$$n = \frac{\beta B^2}{8\pi kT} \tag{4}$$

If this expression is inserted into equation 2, one finds that the fusion power density is given by

$$p_F \sim \left(\frac{\overline{\sigma v}}{T^2}\right) \beta^2 B^4 \tag{5}$$

Here we have three factors which can be manipulated to vary the fusion power density. We shall consider them in turn.

The first factor $(\overline{\sigma v}/T^2)$ represents a physical property of the fuel being used. Fig. 1 shows this quantity as a function of temperature for two fuels, a deuterium-tritium (D-T) mixture, and pure deuterium (D). Note that the quantity is some 100 times higher for the D-T than for the D-D reaction. Since this is a property of the material, we cannot change it; we can only choose our operating temperature. In the case of the D-T reaction, the maximum of the curve is well above the ignition point[1] (the temperature at which the release of fusion energy to charged particles is just sufficient to supply the radiative losses, and is therefore the minimum self-sustaining temperature), so that one would hope to operate at this optimum temperature: 150 million K (degrees Kelvin).

On the other hand, the ignition point for the D-D reaction is approximately 400 million K, so that one would have to operate at or above this higher temperature. It is interesting to note that, even though the reaction cross sections continue to climb with energy for a range well beyond the limits of Fig. 1, the presence of T² in the denominator depresses the maxima of these curves to a 15-kv level (1 electron-volt is equivalent to 11,600 K). In practice one would select an operating temperature as near as possible to the maxima of these curves. The fusion power density would then be proportional to the product $\beta^2 B^4$.

With regard to the second factor β^2, mentioned briefly above, a prime problem of theory and experiment will be to provide the highest possible stable confinement β achievable. In fact, it is this problem that is now receiving world-wide intensive study.

The maximization of the third factor B^4 is a very interesting problem in technology beyond the scope of this paper. The presence of the exponent 4 makes this a particularly attractive approach. Coil design involves two problems which ultimately limit progress in this direction. In qualitative terms, if the device is of small size, achievable cooling rates will limit the designer, while strength considerations will dominate in larger sizes. With present-day technology, field strengths up to the level of 100,000 gausses are feasible. Development to extend this limit by a factor of 3 does not seem out of the question.

Fusion Reactors

Let us now consider a complete machine. Since the fusion power density is proportional to $\beta^2 B^4$, the total power generated by the machine as a result of fusion reactions will be given by

$$P_F = k_1 \beta^2 B^4 r^3 \tag{6}$$

where r is some characteristic length of the machine such as the radius of the vacuum tube containing the plasma, and k_1 is a proportionality constant dependent on all the details of the machine. By far the largest amount of power consumed to operate the machine will almost certainly be that required to operate the magnet, and we assume here that, compared with this quantity, other power requirements of the plant may be neglected. In the absence of magnetic materials, the power required to operate any magnet may be represented by

$$P_M = k_2 \rho B^2 r \tag{7}$$

where ρ is the resistivity of the conductor material, B is the magnetic field strength at any particular point in the system, r is a characteristic length of the device (such as the radius of the internal opening), and k_2 is a proportionality constant depending only on the geometric shape of the electric conductor system used to generate the field.

Vol.
Pt. 1
196

It follows that the power for sale from a fusion power plant is given by

$$P_S = P_F - P_M = k_1\beta^2 B^4 r^3 - k_2 \rho B^2 r \qquad (8)$$

If nature were such that the second of these two terms tended to be very much smaller than the first, the future of fusion power plants would appear quite bright, once physical possibility were proven. Unfortunately this is not the case. Evaluations of k_1 and k_2, on the basis of various assumptions derived from technical considerations, show that these two terms are of the same order of magnitude.[3]

Increasing the Economy of a Fusion Plant

If we are to achieve a practical power plant, P_F must be considerably larger than P_M. If it were only slightly larger, the plant would be circulating a very large amount of power within itself in order to sell a small quantity, and the resulting investment cost per output kilowatt would be impossibly high. In other words, the ratio

$$\frac{P_F}{P_M} = k_3 \frac{\beta^2 B^2 r^2}{\rho} \qquad (9)$$

must be greater than 1 to have any power yield at all, and should be considerably higher (probably a minimum of 2) to be of practical interest.

From equation 9 there are four obvious ways to proceed to improve this ratio. The first is to select the best type of machine. This involves k_3 and β. No results are yet available on this question. Much laboratory experimentation must be done to determine the highest percentage β attainable and to determine which form of machine can deliver the highest k_3.

The second method is to increase the field strength but, as mentioned above, this has technical limitations.

The third method is simply to scale up the size of the reactor. The presence of r^2 in equation 9 implies that if fusion reactors are proved physically possible, they can certainly be made practical if built in large enough sizes. A serious difficulty with this approach is that the cost rapidly increases with increasing size and although the investment cost per kilowatt may be brought down within reason by a unit of large enough size, the output might be too large for any market in the foreseeable future.

The fourth method is to choose a material of lower resistivity than that provided by normal use of copper conductors. Silver is out of the question because of its very high cost for a relatively small improvement in ρ. It has been frequently suggested, however, that refrigeration of conductors be used, to take advantage of the decreasing resistivity with temperature.* Let us compare operating a magnet at ambient temperature T to operating it at a low temperature T_0 with an appropriate refrigerator to remove the heat from the load at T_0 and reject it at ambient temperature T. Let the resistance of the magnet R be

$$R = cT \qquad (10)$$

where T is the absolute temperature, and c a coefficient. Note that c is not a constant but is itself a function of the temperature. Then

$$P_M(T) = I^2 c_T T \qquad (11)$$

and

$$P_M(T_0) = I^2 c_{T_0} T_0 \qquad (12)$$

where I is the current required to generate the magnetic field. If we operate at T_0, we need a refrigerator which will require a power P_R, given by

$$P_R = I^2 c_{T_0} T_0 \left(\frac{T - T_0}{T_0}\right) \frac{1}{\eta} \qquad (13)$$

where η is the coefficient of performance of the refrigerator. The total power required to generate the magnetic field is obtained by adding equations 12 and 13. This is approximately

$$P_M + P_R \cong \frac{1}{\eta} [I^2 c_{T_0} T] \qquad (14)$$

This expression is exact for $\eta = 1$ and only about 3% in error at $\eta = 50\%$.

Compare equation 11 with equation 14. If the coefficient c were a constant and if Carnot refrigerators were available, the total power required would be independent of temperature. In practice, c is almost constant down to temperatures in the cryogenic region. Since practical refrigerators fall short of Carnot performance, it would always waste power to cool the magnet with chilled water or even liquid nitrogen.

On the other hand, in the neighborhood of 20 K and below, the coefficient c drops to very low values compared to its value at higher temperatures for many substances. Post has suggested sodium.[4] Laquer has operated cryogenic magnets of copper for several years.[5] Aluminum is another possibility. This approach may ultimately be of very great importance, particularly since it improves the relative yield P_F/P_M without simultaneously increasing the generated power P_F. Factors of improvement of the order of 10 may be achieved by this method. An unknown factor, however, is the cost required for the extremely large cryogenic refrigerators that would have to be developed for such an installation. In recent years, considerable progress has been made in this direction.

Another, less obvious, approach is cyclic field peaking. Division of the two terms in equation 8 to yield equation 9 implies time independence of all the variables. Suppose the confining field B is periodic in time. The mean value of B^4 over a cycle divided by the mean value of B^2 over the cycle will have different values depending on the waveshape. All other factors remaining constant, the former measures the fusion power liberated and the latter the magnet power consumed.

It is possible to select waveshapes which improve the ratio of P_F/P_M; Fig. 2 presents several examples. To illustrate the potential advantages, in the figure is a tabulation of the mean value of B^4 divided by the mean value of B^2 and normalized to constant magnet power by dividing by the mean value of B^2, i.e., $\overline{B^4}/(\overline{B^2})^2$. For d-c the value is, of course, 1, and the other curves should be compared with this. The waves shown are: simple sine wave, optimized biased sine wave, optimized first and third harmonic mixture, and an optimized biased first and third harmonic mixture. In principle the value of this ratio could be raised to arbitrarily high values by inclusion of higher and higher harmonics to give a more peaked wave. But skin-effect considerations limit the highest usable frequencies, and maximum allowable temperature excursions limit the lowest, so that improvement factors greater than three are unlikely. Since the problem is basically one of increasing a ratio (P_F/P_M) from some value between 1 and 2 to a value in excess of 2, such an improvement could be of great importance.

Combinations of Methods

Clearly the type of machine exhibiting the highest intrinsic economic potential should be chosen for a fusion plant, but high fields, large size, cryogenic cooling, and cyclic field peaking cannot all be combined to obtain an ultimate facility. In fact, almost all combinations of these approaches lead to technical difficulties which would require extensive development for full understanding. As examples, the following may be mentioned.

---

* We are discussing only normal conductivity here; the interesting idea of using superconductivity for technical apparatus is a different topic.

The combination of high fields with large sizes leads to the problem of structural strength, mentioned earlier. If high fields are combined with cryogenic cooling, a phenomenon known as the magnetoresistive effect causes trouble. The resistance of a conductor rises as the magnetic field through it is increased. At room temperatures, this is a very unimportant effect, but at cryogenic temperatures it can be the major component of the resistance. If peaking is to be combined with high fields and large volumes, one finds that thousands of volts per turn may be required in the magnet to cycle the large fluxes involved, and insulation becomes a serious problem. Finally, the combination of cryogenic cooling and cyclic field peaking seems out of the question because of the skin-effect problems. The skin depth decreases with the square root of the resistivity, and becomes so small at cryogenic temperatures (0.01 inch at 60 cycles) that one would have to wind large magnets with wire so fine that fabrication and insulation would be impossible. The two most promising methods seem to be cryogenic d-c magnets and ambient temperature peaked-field magnets.

This paper has considered only that class of machine in which an externally generated magnetic field is used. Thus all pinch machines are excluded, but included are the stellarator,[6] magnetic mirror,[7] astron,[8] miscellaneous cusped geometries,[9,10] the interesting form of stellarator suggested by Lehnert,[11] and others. It is too early to predict which of these, if any, will prove to be potentially an economic reactor, and far too early to make any good cost estimates of a resulting plant.

References

1. Controlled Fusion Research—An Application of the Physics of High Temperature Plasmas, **R. F. Post.** *Reviews of Modern Physics*, New York, N. Y., vol. 28, 1956, p. 338.

2. Some Stable Hydromagnetic Equilibria, **J. Johnson, C. Oberman, R. Kulsrud, E. Frieman.** *Physics of Fluids*, New York, N. Y., vol. 1, 1958, p. 81.

3. Problems of the Stellarator as a Useful Power Source, **L. Spitzer, D. Grove, W. Johnson, L. Tonks, W. Westendorp.** *Report no. NYO-6047*, Atomic Energy Commission, Washington, D. C., 1954.

4. Air Core Cryogenic Magnet Coils for Fusion Research and High Energy Physics Applications, **R. F. Post, C. E. Taylor.** "Advances in Cryogenic Engineering" (book), Plenum Press, Inc., New York, N. Y., vol. 5, 1959.

5. Cryogenic Electromagnets. I.—Feasibility Study, **H. L. Laquer, E. F. Hammel.** *Review of Scientific Instruments*, New York, N. Y., vol. 28, 1957, p. 875.

6. The Stellarator Concept, **L. Spitzer, Jr.** *Physics of Fluids*, vol. 1, no. 4, July-Aug. 1958, pp. 253–64.

7. Summary of UCRL Pyrotron (Mirror Machine) Program, **R. F. Post.** *Proceedings*, Second U. N. International Conference on the Peaceful Uses of Atomic Energy, Geneva, Switzerland, vol. 32, 1958, pp. 245–56.

8. Astron Thermonuclear Reactor, **N. Christofilos.** *Ibid.*, pp. 279–90.

9. Picket Fence, **J. L. Tuck.** *Report no. WASH-184*, Atomic Energy Commission, Oct. 26–27, 1954, pp. 77–86.

10. Cusped Geometries, **J. Berkowitz, K. O. Friedrichs, H. Goertzel, H. Grad, J. Killeen, E. Rubin.** *Proceedings*, Second U. N. International Conference on the Peaceful Uses of Atomic Energy, vol. 31, 1958, pp. 171–76.

11. On the Confinement of Charged Particles in a Magnetic Field, **B. Lehnert.** *Journal of Nuclear Energy*, New York, N. Y., vol. 1, pt. C, 1959, p. 40.

LATE DISCUSSION

Discussion of "Properties of Octafluorocyclobutane, a Dielectric Gas," by F. W. Blodgett, published in Communication and Electronics, March 1959, pages 63–66.

S. Yamakura (Kanagawa, Japan): I would like to comment on the interesting paper describing a new dielectric gas, octafluorocyclobutane (OFCB) by Dr. Blodgett. In the near future, I plan to carry out some fundamental experiments and applications.

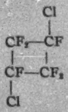
We expect dielectric gases to be chemically inert and nontoxic, and to have high dielectric strength. Most fluorogases are chemically stable, e.g., fluorocarbons and sulfur hexafluoride.

OFCB is highly stable, but we suppose that at high temperatures it produces some products of decomposition. The decomposition temperature, for long-term use, is higher than that for sulfur hexafluoride, but the gas is slowly decomposed in a corona discharge, and the chemical properties change.

Many engineers hope that the gas will have high dielectric strength in uniform and nonuniform fields. OFCB has a high dielectric strength compared with sulfur hexafluoride. Of course, these gases tend to be electronegative in the molecular state. During the past half-century the mechanism of dielectric breakdown of gases has been widely discussed, but no entirely satisfactory theory has been developed.

Fig. 2 shows that the corona-starting voltage of OFCB is about five times that of air. This is important data for design engineers. The breakdown voltage as a function of pressure is shown in Fig. 4(A) as a straight line for a uniform field, but random variations probably occur. I am now experimenting with sulfur hexafluoride imported from England. At low pressures, the random variation is 20%, so that I assume the author's data may have been averaged. In a nonuniform field, the maximum dielectric strength is shown at relatively low pressures; see Fig. 4(B). I don't know just how space charge might distort these values, but sulfur hexafluoride appears superior to OFCB.

Hopefully, two chlorine atoms could be introduced symmetrically into the molecule to give a higher dielectric strength than that of OFCB. The boiling point would be higher, but the dipole moment would still be zero.



The evaluation of physiological hazards is difficult for new chemical products which are still under development. When they are not completely pure, any observed physiological effects may be caused either by the impurities or by the pure material itself. Dr. Blodgett points out that the toxicity of OFCB is very low, but we hope that this will be investigated further.

F. W. Blodgett: Mr. Yamakura has raised several interesting points about the new dielectric gases. The extent of decomposition of OFCB at high temperatures, in a corona discharge, or in an arc, is difficult to evaluate. Some surfaces are more catalytically active than others. The intensity of the corona or the arc in relation to the amount of OFCB is usually unknown. New analytical tests are being developed to detect and identify the decomposition products. The chemical breakdown of the gas, and the final products isolated, may eventually contribute to our understanding of the electrical breakdown process.

Random variations of 10% to 20% in corona-starting voltage and breakdown voltage data do occur. It is important to control factors such as the finish and cleanliness of the electrodes used, the illumination of the electrode surfaces, and the static charges which tend to accumulate on the surfaces of solid insulation near the electrodes. The data shown in the figures are average values.

Fig. 4(B) shows the complexity of the breakdown voltage as a function of pressure in a particular nonuniform field. OFCB has a higher breakdown voltage than sulfur hexafluoride, but both maxima appear at low pressures. The maxima may be moved to higher pressures by diluting the electronegative gases with nitrogen, as shown. Some of the lower molecular-weight fluorocarbon gases show broader but lower maxima versus OFCB (see reference 1 of the paper).

The symmetrical trans-1,3-dichlorohexafluorocyclobutane suggested by Mr. Yamakura would indeed have a zero dipole moment, a higher boiling point (about 60 C), and a higher dielectric strength as compared to OFCB. Unfortunately, it would be less stable chemically, and probably anesthetic or slightly toxic. The presence of the chlorine atoms would lead to corrosion of metals and deterioration of cellulosic insulation if any corona or arcing should occur in a system containing traces of moisture.

OFCB is now made in very high purity, and the pure material was found to be essentially inert physiologically. Animals showed no anesthesia or other ill effects, and no tissue pathology at autopsy, after exposure to 80% OFCB and 20% oxygen (by volume) for 6 hours, or after 90 6-hour exposures to 10% OFCB in air containing 20% oxygen.

As indicated above, however, slight decomposition of OFCB, with the formation of some toxic products, does occur at high temperatures or after electrical breakdown. Although OFCB is nontoxic as supplied, it is necessary to avoid breathing gas after it has been exposed to fire, arc, or corona. We expect to develop suitable adsorbents that can be placed in the dielectric gas enclosure to remove any decomposition products formed.